GENOMICS OF COELIAC DISEASE

Molecular Signatures of the Pathogenesis

ISBN-10: 90-9021312-0
ISBN-13: 978-90-9021312-5

Cover, layout and print: Budde Grafimedia bv, Nieuwegein

The work presented in this thesis was performed at the DBG - Department of Medical Genetics of the University Medical Center Utrecht, and financially supported by the Dutch Digestive Disease Foundation (MLDS) grants WS 00-13 and WS 97-44, the Netherlands Organization for Medical and Health Research (NWO) grants 902-22-094 and 912-02-028, and by the Celiac Disease Consortium, an Innovative Cluster approved by the Netherlands Genomics Initiative and partially funded by the Dutch Government (BSIK03009).

The publication of this thesis was facilitated by a financial contribution from the Department of Medical Genetics and generous donations from the Dutch Digestive Disease Foundation (MLDS); the Section of Experimental Gastroenterology (SGE) of the Netherlands Society of Gastroenterology (NVGE); the JE Juriaanse Stichting; and the Dr Ir JHJ van der Laar Stichting.

GENOMICS OF COELIAC DISEASE

Molecular Signatures of the Pathogenesis

GENOMICS VAN COELIAKIE

Moleculaire Kenmerken van de Pathogenese

(met een inleiding en samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. W.H. Gispen,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
op dinsdag 28 november 2006 des middags te 2.30 uur

door
Mari Cornelis Wapenaar
geboren op 14 juli 1958 te Vlaardingen

Promotor: Prof. dr. C. Wijmenga

Eén koorengraan,
wie zal 't verstaan?
behelst tienduizend schooven:
gelooft gij niet
als 't geen gij ziet,
hoe zult gij dit gelooven?
Guido Gezelle (1830 - 1899)

You cannot depend on your eyes when your imagination is out of focus
Mark Twain (1835 – 1910)

Ter nagedachtenis aan mijn ouders

In loving memory of Lisa:
still so much to accomplish
so much to live for
so much to love

CONTENTS

PREFACE

Definition

"Coeliac disease; a malabsorption syndrome precipitated by ingestion of gluten-containing foods and marked by diarrhea with bulky, frothy, fatty fetid stools, abdominal distention, weight loss, asthenia, deficiency of vitamins B, D, and K, and electrolyte depletion" (from 24th Edition of Dorland's Pocket Medical Dictionary, DM Anderson, WB Saunders Comp, Philadelphia). *When I started this research six years ago, one of my first encounters with coeliac disease was this rather graphic description that, admittedly, features a certain poetic beauty.*

Aim

The aim of this thesis was to gain insight into the pathogenesis of coeliac disease by means of a genomics approach. More explicitly, we carried out integrated genetic analysis and gene expression studies to identify, and gain better understanding of the genes and mechanisms behind the pathology and etiology of coeliac disease.

Outline

In the introductory chapters 1 and 2 (in Dutch with English summary), coeliac disease is presented as a gluten-intolerance and complex genetic disorder. Information is provided on the currently identified susceptibility genes and loci. Genetic strategies for the identification of new at-risk genes are discussed, and include sib-pair linkage analysis and genetic association. Gene expression studies are discussed as a tool for gaining insight into the pathology and for identifying genes and molecular pathways to aid the genetic analysis. Special emphasis is placed on the new high-throughput screening technologies and their application for gene expression studies is described in section II. In chapter 3 we deal with the issue of heterogeneity in patients' tissue samples and to what extent this has an impact on gene expression studies. Much to our surprise, we observed that gene expression mosaicism was always present, regardless of whether there was tissue patchiness or not. This led us to propose a model to explain the emergence of tissue patchiness from the existing gene expression mosaicism. Our first genome-wide expression study using cDNA-microarrays is described in chapter 4. In these experiments we compared patients with flat mucosa to controls and identified several dynamic processes, including proliferation and differentiation. The differentiation-stimulating and proliferation-attenuating gene *TM4SF4* that emerged in this study also appears in chapters 3 and 5. An alternative strategy, using oligonucleotide-microarrays and patients in various degrees of remission, is described in chapter 5. Here we demonstrated that incomplete differentiation of enterocytes causes the impairment of multiple nutrient-absorption functions of the intestine, which can be related to the diverse clinical features seen in coeliac

disease patients. In addition, we observed the loss of intestinal detoxification, a feature that has also been reported for inflammatory bowel disease.

In section III we depart from the genome-wide approach and focus on a specific molecular pathway, the tight junction gene network. Epithelial tight junctions seal the luminal environment from the internal circulation and a breach in this barrier could contribute to the pathogenesis of coeliac disease. Chapter 6 is a descriptive study of the transcriptional activity of the genes comprising the tight junction network and how they change under the influence of coeliac disease. The same gene network was investigated in a genetic association study that made use of single nucleotide polymorphisms and linkage disequilibrium maps. This yielded two candidate genes that not only contribute to coeliac disease susceptibility, but also enhance the risk of developing inflammatory bowel disease.

Section IV deals with the genomics of candidate genes related to the immune defense. Interferon-gamma *(IFNG)* is the major pro-inflammatory cytokine that drives the adaptive immune response in coeliac disease and evokes the tissue remodeling that leads to a flat mucosa. In chapter 8 we describe the relationship between *IFNG* expression and the extent of mucosal remodeling in patients. Two types of genetic analyses, using a microsatellite from the first intron of this gene, failed to reveal any association in Dutch patients. However, by using genetic analysis with single nucleotide polymorphism markers we were able to demonstrate weak association, as described in chapter 9. The *SPINK* gene family plays an important role in tissue preservation and bacterial containment, and could be considered part of the innate immune system. Particularly the interesting location in the genome of individual members of this gene family led us to decide to perform a genetic association test. However, no association could be observed in Dutch coeliac disease patients, as reported in chapter 10. Finally, in the Discussion, I attempt to unify our findings, and those of other workers, into a model for the pathogenesis of coeliac disease.

SECTION I

INTRODUCTION

Chapter 1

A combined genetics and genomics approach to unravelling molecular pathways in coeliac disease

MC Wapenaar and C Wijmenga

*Novartis Found Symp 2005; **267**:113-34*

(reproduced with permission from John Wiley & Sons Limited)

A combined genetics and genomics approach to unravelling molecular pathways in coeliac disease

Martin C. Wapenaar and Cisca Wijmenga[1]

Complex Genetics Group, Department of Biomedical Genetics, University Medical Centre Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands

Abstract. Coeliac disease (CD) is a complex, inflammatory disorder of the small intestine induced by gluten. It is common and has a prevalence of ~1:200 in Western populations. A major known susceptibility locus for CD is the *HLA-DQ* locus. However, the genetic contribution of this region is limited to ~40%, so non-*HLA* genes must also be involved in the disease aetiology. Genetic studies have so far identified multiple loci that may potentially be involved in disease aetiology, although the majority of these loci are expected to point to genes with a small effect. A major CD locus on chromosome 19 was recently identified in the Dutch population. Interestingly, there is some marked overlap when comparing genetic linkage studies conducted in different autoimmune disorders, suggesting that common pathways contribute to these diseases. This knowledge may eventually help in identifying some of the disease genes. To identify the true disease-causing genes, linkage analysis needs to be followed by genetic association. Because of the nature of the probable mutations, it is to be expected that the investigation of gene expression data can assist in selecting candidate genes from linkage regions. Furthermore, expression data will point to the molecular pathways involved in the disease pathogenesis.

2005 The genetics of autoimmunity. Wiley, Chichester (Novartis Foundation Symposium 267) p 113–144

Coeliac disease (CD; MIM 212750) is a chronic disorder caused by an inflammatory response to gluten, a dietary product present in wheat, barley and rye. Ingestion of gluten by CD patients leads to flattening of the duodenal mucosa, a gradual process classified according to Marsh (Fig. 1). CD occurs largely in Caucasians. The disease is less common in men, with a male to female ratio of 1:3. Recent studies revealed that the prevalence of CD is 1 in 100–300, which is much higher than previously thought. Although CD is one of the most

[1]This paper was presented at the symposium by Cisca Wijmenga to whom all correspondence should be addressed.

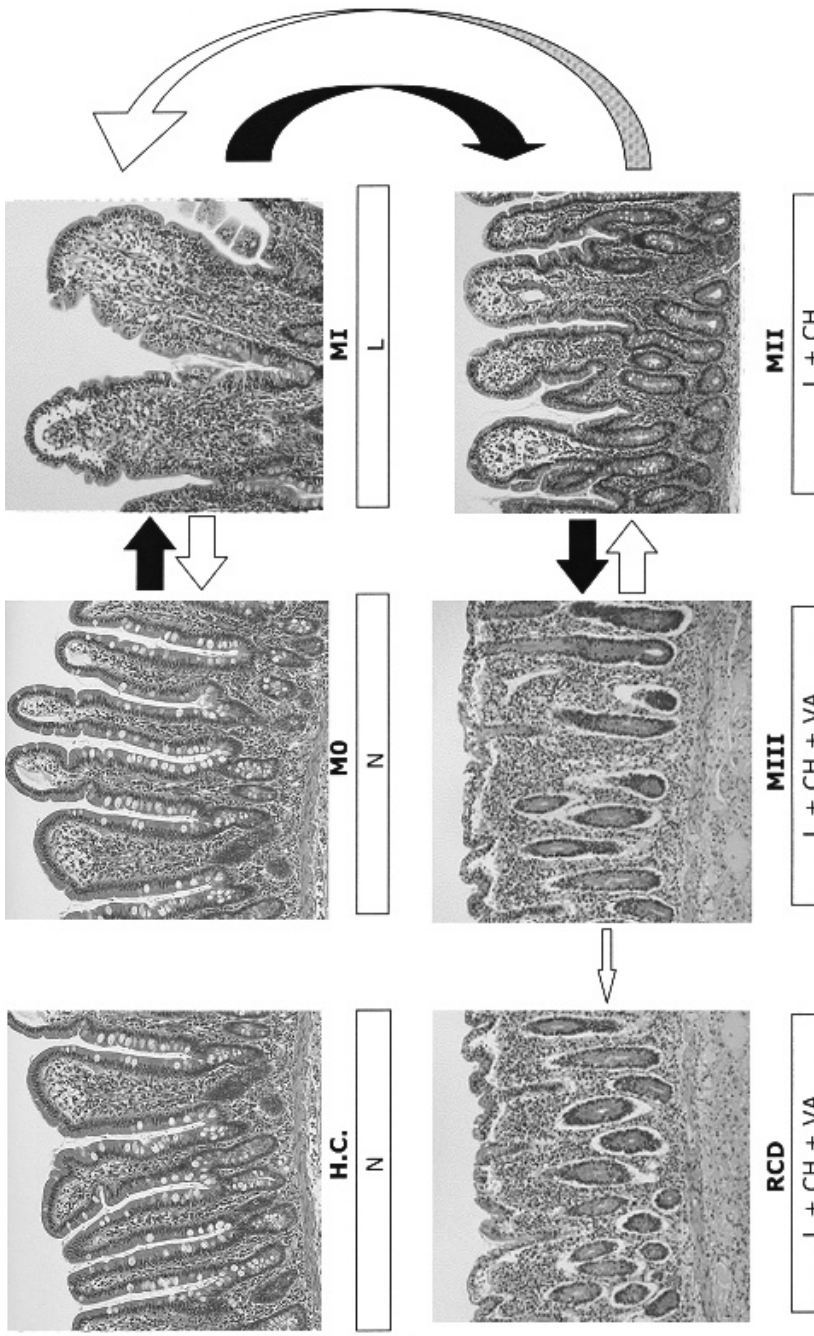


FIG. 1. Ingestion of gluten (black arrows) by CD patients (M0) results in lesions of the proximal small intestine. The range of abnormalities can be classified according to the Marsh classification. Marsh I (MI) comprises normal mucosal architecture with marked lymphocytosis (L). MII includes intraepithelial lymphocytes and crypt hyperplasia (CH) with branching and elongation of crypts. MIII comprises additional villous atrophy (VA). Upon treatment with a gluten-free diet (white arrows), the majority of patients undergo complete remission (M0), comparable to healthy controls (HC). A small percentage of patients become refractory (RCD) to the diet.

common forms of food intolerance in the world, approximately 85% of affected individuals go unrecognized.

Clinical presentation

The clinical presentation of CD comprises a wide spectrum of symptoms, most of them related to the malabsorption of nutrients from food. Typical symptoms of childhood CD include chronic diarrhoea, abdominal distension and a failure to thrive. Recurrent aphthous lesions in the mouth, dental enamel defects, fatigue, or an isolated iron-deficiency (anaemia) may also be manifestations of CD, both in children and adults. Other manifestations of CD may be dermatitis herpetiformis and osteoporosis. Moreover, infertility or recurrent abortions have also been observed in women with CD. Besides the paucisymptomatic presentation, some patients are even asymptomatic. Strict adherence to a gluten-free diet results in complete restoration of the small intestine and disappearance of the clinical symptoms.

CD is diagnosed in $\sim$4.5% of patients with type 1 diabetes mellitus (T1D) and autoimmune thyroid disease (ATD). Similarly, approximately 5% of patients with ATD are found to be positive for CD. This co-morbidity may suggest a common disease aetiology.

In recent years, serological tests have become available to screen for CD. However, the standard for CD diagnosis is still an intestinal biopsy sampling, an invasive procedure that requires general anaesthesia.

CD is a multifactorial disorder

CD is a typical example of a multifactorial disorder, i.e. a disease caused by the combined action of several genes and environmental factors. A recent Italian twin study found higher concordance rates for CD in monozygous (MZ) twins (86%) than in dizygous (DZ) twins (20%), indicating a strong genetic component in the development of CD (Greco et al 2002). CD aggregates in families with an approximate 10% recurrence risk for siblings of CD patients. So, based on a population prevalence of 0.5%, the sibling relative risk (λ_s) for development of CD is 20. There is, however, no Mendelian inheritance of the disease in families, suggesting that the genetic predisposition is derived from more than one gene. An important role for environmental factors is evident from the less than 100% concordance rate in MZ twins. So far, gluten is the only proven risk factor; it is also the major environmental risk factor for CD.

A strong association between HLA-DW3 and CD was identified as far back as 1976 (Keuning et al 1976). We now know that the primary association is with the genes encoding the HLA-DQα1 and HLA-DQβ1 peptides of the HLA-DQ2

heterodimer, localized in the major histocompatibility complex (MHC) on chromosome region 6p21.3. The HLA-DQ2 molecule, encoded by the *HLA-DQA1*05* and *HLA-DQB1*02* alleles in either the *cis* or the *trans* configuration, is expressed by more than 90% of CD patients. This is in strong contrast to the frequency of HLA-DQ2 carriers in the general population, which is 20–30%. A causal role for HLA-DQ2 has been demonstrated as this molecule can present the gluten-derived peptides to T cells.

Although HLA-DQ2 seems to be necessary for CD development, it is certainly not sufficient in itself. HLA-DQ2 alone can explain some 40% of the genetic variation underlying CD, implying that there are other non-HLA genes which are also important determinants of disease aetiology. A full understanding of the molecular events involved in the disease pathogenesis involves identification of the full repertoire of disease susceptibility genes.

Genetic studies conducted in CD

Both family-based linkage studies and population-based association studies have been performed to identify non-HLA genes (Fig. 2).

Genomewide linkage studies

Genomewide linkage studies have identified a number of putative loci across different populations (table 1), apart from the well-established HLA association. A locus on 5q31–q33 (*CELIAC2*) has been repeatedly identified in various genomewide linkage screens as well as in a meta- and mega-analysis by a large EU consortium, albeit with only suggestive evidence in most of the studies, implying a locus of modest effect. Interestingly, this same region has been implicated in inflammatory bowel disease (IBD) (IBD5; Rioux et al 2001), ATD (Sakai et al 2001), asthma (Xu et al 2001) and rheumatoid arthritis (RA) (Tokuhiro et al 2003), suggesting common inflammation and/or autoimmunity genes. The 5q31–q33 region contains a number of cytokine genes, none of which has so far shown association to CD (Ryan et al 2004) or to any of the other diseases in this region. Furthermore, the *CELIAC3* locus encompassing the *cytotoxic T-lymphocyte-associated antigen 4* (*CTLA4*) gene on chromosome 2q33 has been identified by both linkage and association studies in several different populations. However, this locus is considered to confer only a very modest risk for susceptibility to CD pathogenesis. *CTLA4* is an interesting functional candidate gene for T cell-mediated disorders. This is also in concordance with the identification of strong association between single nucleotide polymorphisms (SNPs) in *CTLA4* and type 1 diabetes (T1D), Graves' disease (Ueda et al 2003) and asthma. More recently a significant locus on chromosome 19 (*CELIAC4*) was

identified in the Dutch population (van Belzen et al 2003), which awaits formal replication. The EU meta-analysis also suggested this locus, although the lod score was modest (Babron et al 2003). Surprisingly, this same region showed suggestive evidence for linkage in a recent meta-analysis of Crohn's disease (van Heel et al 2004). Ten additional genetic loci have been identified for CD that await formal replication (Table 1). Amongst these, the 9p21–p13 locus is of interest since this region has now been implicated in three different studies. Apart from the suggestive linkage with a lod score of 2.61 found by van Belzen et al (2004), studies in the Swedish/Norwegian (Naluai et al 2001) and the Finnish population (Liu et al 2002) showed lod scores of 1.78 and 1.11, respectively, with markers from the same region. Another interesting locus is on 6q21 (van Belzen et al 2003) since this region may also contain loci for T1D (*IDDM15*; Cox et al 2001), multiple sclerosis (MS) (Akesson et al 2002) and RA (Jawaheer et al 2003). It cannot be excluded that some of the CD genes actually predispose to a general susceptibility to autoimmunity and/or inflammation. Given the high co-morbidity of CD with other autoimmune disorders, common pathways may be involved in the destruction of the target tissues. Such a hypothesis, alluded to as the 'common variants/multiple disease' model, implies that many disease genes may not be disease-specific. Interestingly, this hypothesis might be correct as a comparison of a large number of linkage screens for autoimmune disorders showed substantial non-random clustering (Becker et al 1998).

So far 10 different genomewide linkage scans have been performed for CD. There is growing evidence that at least some of the regions identified will contain true susceptibility genes. Why has it been so difficult to identify these loci and why are there so many differences found by the various studies? There are a number of remarks we can make:

- In general, each of the samples studied is relatively small, ranging from a single large family to 102 families consisting only of affected pairs of relatives. Hence, each study has limited power in detecting new loci and even less power to replicate findings obtained in other studies.
- A large number of different populations have been studied, ranging from genetically homogeneous populations such as the Finns, to rather outbred populations such as the Americans. Therefore, it cannot be excluded that population-based differences are present in the magnitude to which the different loci contribute to disease susceptibility. This is actually already evident for both the *CELIAC2* and *CELIAC3* loci. For example, the Italian population shows strong linkage to *CELIAC2* on 5q31–q33 whereas most other populations show only modest lod scores that usually do not exceed the level of suggestive evidence for linkage.

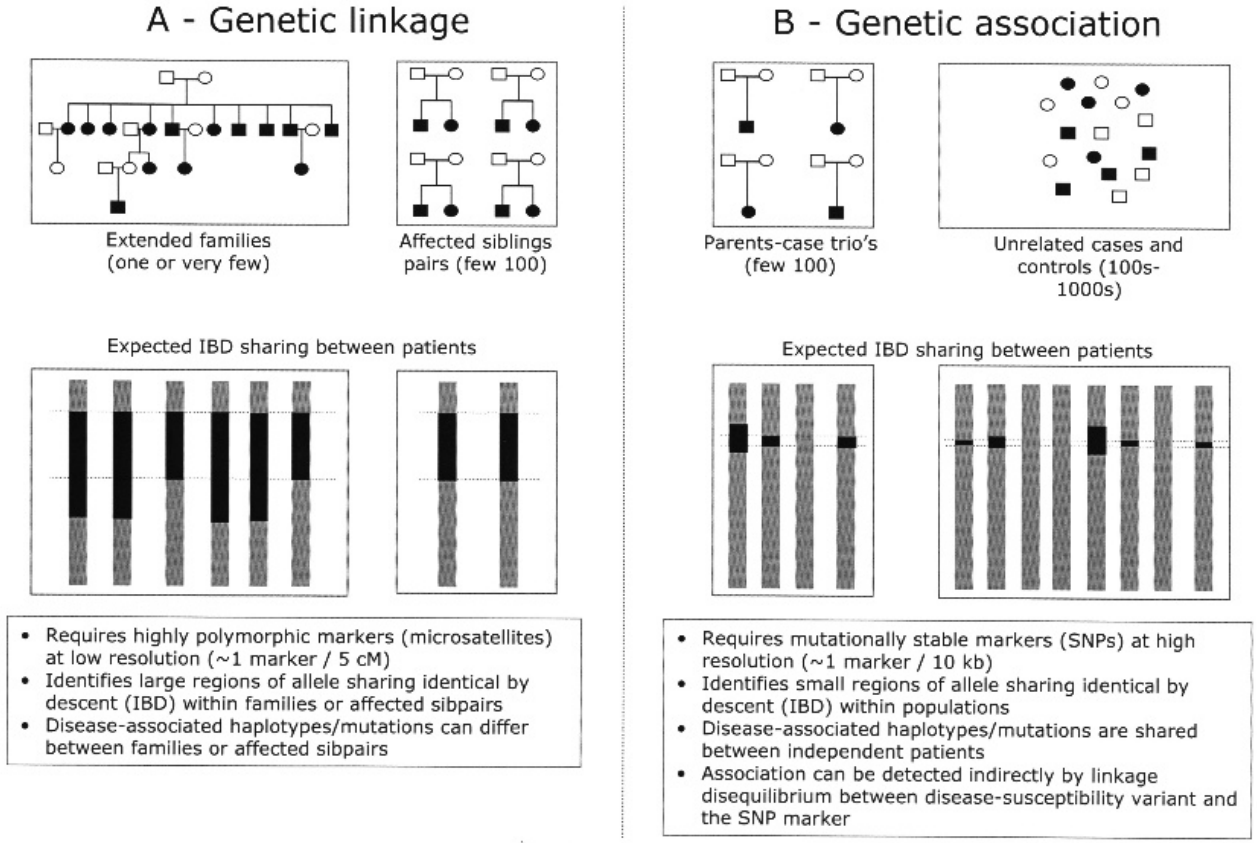


FIG. 2. Two common genetic approaches can be used to find genetic variants (alleles) involved in the aetiology of CD: genomewide scans using highly polymorphic markers to identify chromosome regions harbouring disease-risk genes, and association studies between the disease phenotype and sequence variants in defined candidate genes. (A) Linkage studies in complex diseases are mainly being performed using affected pairs of siblings by searching for regions of excess allele sharing between affected individuals from one family (depicted in black). Siblings share 50% of their DNA on average, so a large number of sibpairs is required to detect linkage. The use of extended families is less common. (B) Association studies make predictions about the nature of the genetic variation underlying the disease (the CD/CV model). As a consequence, the identity-by-descent (IBD) region shared between independent patients is expected to be small (depicted in black). Association studies require many patients and can be both family-based (using the transmission disequilibrium test, TDT) or population-based (using case-control design). TDT can be used to avoid effects due to population stratification.

TABLE 1 Genome-wide linkage studies performed in CD

CD locus	*Chromosomal location*	*Genes and tested candidates*	*Locus relative risk*	*Mapping method*[a]	*Significance*[b]	*Population*	*References*
CELIAC2	5q31-q33			GWS, SR	Sug	Italian	Greco et al 1998, 2001
				Meta/ mega-analysis	Sug	EU CD consortium[c]	Babron et al 2003
CELIAC3	2q33	CTLA4, ICOS	Very modest	SR	Sug	Finnish	Holopainen et al 2001
				GWS-ASP	Sug	Swedish/Norwegian	Naluai et al 2001
CELIAC4	19p13.1		2.6	GWS-ASP	Sig	Dutch	Van Belzen et al 2003
–	4p15			GWS-ASP	Sug	Finnish	Liu et al 2002
–	6p23			GWS-ASP	Sig	Irish	Zhong et al 1996
–	6q21			GWS-ASP	Sug	Dutch	Van Belzen et al 2003
–	6q25.3			4-generation family	Sug	Dutch	Van Belzen et al 2004
–	9p21-p13			4-generation family	Sug	Dutch	Van Belzen et al 2004
–	10p			GWS-ASP	Sug	Finnish	Rioux et al 2004
–	11p11			GWS-ASP	Sig	Irish	Zhong et al 1996
–	15q11-q13			GWS-IP	Sig	Finnish	Woolley et al 2002
–	15q26			GWS-ASP	Sug	Irish	Zhong et al 1996
–	22 cen			GWS-ASP	Sug	Irish	Zhong et al 1996

[a]GA genetic association, GWS genomewide scan, MMA meta/mega-analysis, SR selected regions based on previous genomewide scan data, GWS in ASP genomewide scan in affected sibpairs; GWS in IP genomewide scan in isolated populations.
[b]Sig = significant linkage: $P < 2.2\times10^{-5}$ (lod score > 3.6); Sug = suggestive linkage: $P < 7.4\times10^{-4}$ (lod score > 1.9). According to Lander and Kruglyak. The level of significance was based on multipoint values if available.
[c]The EU consortium consists of CD families from Finland, Sweden, Norway, UK, France and Italy.

- Although the gold standard for diagnosing CD is an intestinal biopsy, it is also clear that not every patient included in the published studies has been subjected to the same stringent diagnostic criteria. A study in the Netherlands in which all initial biopsies were re-evaluated by the same pathologist who specialized in CD, revealed some 20% misclassified samples, which were then excluded from the genomewide screen (van Belzen et al 2003). Inclusion of these samples might have seriously jeopardized finding significant linkage.

Functional candidate gene studies

Based on our limited knowledge of the disease process underlying CD, a number of pathways have been implicated, and hence genes from these pathways have been tested for their possible involvement in CD by genetic association studies. The most studied genes include those from immunological and inflammatory response pathways, including *CTLA4*, interferon gamma (*IFNγ*), interleukin 12B (*IL12B*) and interleukin 6 (*IL6*). In addition, transglutaminase 2 (*TG2*) has been studied because of its role in gluten modification. So far, none of these studies have shown evidence for involvement in CD pathogenesis, apart from inconsistent results obtained for *CTLA4* (see Table 2), both with respect to genetic association with SNPs in *CTLA4* as well as with markers covering the *CELIAC2* region (for an overview see Holopainen et al 2004). These results might suggest that the true susceptibility CD gene from this region has not yet been identified since rather strong association was observed for a region 2 Mb away from *CTLA4*. For the other functional candidate genes that were tested, it is probably still too early to truly exclude them since they might not always have been studied in a proper way. As the risk associated with most of the functional candidate genes is probably extremely modest — since the chromosomal regions to which they map usually do not show up in linkage studies — most studies might have been underpowered. Moreover, the incorrect gene variant might have been studied due to a lack of knowledge on the linkage disequilibrium structure of the gene.

How to go from a linkage region to a disease gene?

There are a number of promising linkage regions to investigate further for true CD susceptibility genes, including *CELIAC2* (5q31–q33), *CELIAC4* (19p13.1), and the regions on 9p21–p13 and 6q21. Unfortunately, each of these regions is several megabases in size and contains at least 100 different genes. None of the regions harbour obvious candidate genes, requiring the entire candidate regions to be scrutinized for the true disease susceptibility gene. The general strategy is to perform association studies across the entire region with SNPs tagged towards all the haplotypes spanning the locus. Once positive association is observed this

TABLE 2 Association studies of the *CELIAC 3* region encompassing *CD28/CTLA4/ICOS*

Population	*Associated locus*	*P-value*	*Map location (Ensemble v20.34c.1)*	*Distance to CTLA4*	*Reference*
French	*CTLA4 (+49*A)*	< 0.0001	204941 kb	–	Popat et al 2002a[a]
Finnish	*D2S116*	0.0001	201870 kb	3071 kb	Popat et al 2002a[a]
Swedish/	*D2S2392*	0.037	199913 kb	5028 kb	Popat et al 2002a[a]
Norwegian	*D2S2214*	0.044	202933 kb	2008 kb	
	*CTLA4 (+49*A)*	0.012	204941 kb	–	
UK	*D2S2214*	0.007	202933 kb	2008 kb	King et al 2002
Italian	*CTLA4 (+49*A)*	0.003	204941 kb	–	Mora et al 2003
Swedish	*CTLA4 (+49*A)*	0.02	204941 kb	–	Popat et al 2002b
Finnish	*ICOS haplotype*	0.0006	205011 kb	70 kb	Haimila et al 2004
Meta-analysis[b]	*D2S116*	0.0006	201870 kb	3071 kb	Popat et al 2002a
	D2S2214	0.0014	202933 kb	2008 kb	
	*CTLA4 (+49*A)*	0.019	204941 kb	–	
Meta-analysis	*D2S72*	0.017	204962 kb	21 kb	Holopainen et al
EU-CD	*D2S2214*	< 0.05	202933 kb	2008 kb	2204
consortium[c]					

[a]This reference summarizes seven different studies.
[b]The meta-analysis consist of CD families from Italy, Tunisia, Finland, Sweden, Norway, the Netherlands, UK, France and northern Europe.
[c]The EU consortium consists of CD families from Finland, Sweden, Norway, UK, France and Italy.

should be replicated in independent populations that also show linkage to the same region. This strategy however assumes that the underlying disease mutation or variation is a common variant — this is generally alluded to as the common variant/common disease (CV/CD) theory. There is ample evidence that this CV/CD theory is indeed correct. Well-known examples include the factor V Leiden mutation in relation to venous thrombosis, and the *ApoE4* allele in Alzheimer's disease. There is, however, also evidence to the contrary. Many different mutations have been identified in the *NOD2* gene that plays a role in the aetiology of IBD (Hugot et al 2001).

In Mendelian diseases the majority (59%) of mutations involve in-frame amino acid substitutions. For non-Mendelian multifactorial disorders we might however expect a quite different distribution of the types of mutations, such as regulatory mutations — these account for only 0.8% of the total in Mendelian disorders. The net effect of such mutations might be a change in the level of gene expression, as described recently for *CTLA4* (Ueda et al 2003). Another example includes

psoriasis, where a disease-associated SNP leads to the loss of a RUNX1 binding site and might be responsible for the three- to fivefold down-regulation of *SLCA3R1* in activated T cells (Helms et al 2003). Interestingly, regulation of *SLC22A4* expression by RUNX1 is associated with susceptibility to RA (Tokuhiro et al 2003). These examples suggest that studies aimed at genomewide expression profiling might actually aid in identifying those disease susceptibility genes that are associated with changes in transcription levels.

Gene expression profiling

Identification of genes with altered expression due to their involvement in disease aetiology and/or pathology is facilitated by microarray hybridizations or quantitative real time polymerase chain reaction (qRT-PCR). Microarray hybridization allows genomewide probing of transcriptional activity, although the currently available arrays are incomplete with respect to genome coverage. qRT-PCR is superior in sensitivity and accuracy compared to microarrays, but allows only a smaller number of genes to be tested in a single assay (10–400 genes). The type of biological sample selected for testing may also have a significant impact on the expression profile obtained. Tissue samples provide an overall insight into expression dynamics but transcriptional changes in a specific cell type may be too diluted and go undetected in this cumulative expression profile. Although specific cell types can be isolated by cell sorting or laser capture microdissection, they may lead to a biased approach that fails to observe relevant changes in cell types not tested for. For CD, two microarray studies on whole mucosal biopsy samples have been published (Diosdado et al 2004, Juuti-Uusitalo et al 2004). While both studies identified gene sets that showed little overlapping, when combined they pointed towards changes in similar processes: activated Th1 response, enhanced cell proliferation, reduced epithelial differentiation, recruitment of $\gamma\delta$T cells, B cells and macrophages, and the lack of changes in matrix metalloproteinases (Table 3). The unexpected finding of up-regulated genes involved in cholesterol and lipid metabolism (Diosdado et al 2004) was previously also observed in a microarray study on murine $\gamma\delta$T intraepithelial lymphocytes (Fahrer et al 2001), thereby revealing the increase of this cell type in the CD lesion. This also demonstrates the added power of comparing expression studies on whole tissue samples with their constituent cell types or a subset of them.

The ability to distinguish the healthy from the CD mucosa based on the expression profile opens the possibility of assembling a panel of marker genes for molecular phenotyping to assist in diagnosis and prognosis, similar to that for breast cancer, for example. Further dissemination of the CD pathological processes can be obtained by profiling biopsy samples from patients on a gluten-

TABLE 3 Microarray studies performed on CD duodenal biopsy samples

Reference	*Diosdado et al 2004*	*Juuti-Uusitalo et al 2004*	*Diosdado et al ongoing work*
No. of genes on microarray	19200	5184	21000
Patient groups tested[a]	HC (7), CD MIII UT (7), CD MIII T (4), RCD MIII T (4)	HC (4), CD MIII UT (4), CD M0 T (4)	HC (21), MO T (11), MI T (8), MII T (10), CD MIII T (3), CD MIII UT (12), RCD MIII T (4)
No. of diff. expressed genes	229 (109: MIII *vs* HC; 120 CD MIII T *vs* CD MIII UT)	263 (156 CD MIII UT *vs* HC; 60 CD M0 T *vs* HC)	169 (over all categories)
Statistical analysis	Welch *t*-test	Manual inspection	MAANOVA
Pathways involved			
Th1 response	+	+	+
Role for $\gamma\delta$T cells	+	—	+
B cell maturation	—	+	—
Role for macrophages	+	—	—
Cell proliferation	+	—	+
Cell differentiation	+	+	+
Genes under linkage peaks	$P < 0.05$	Manual inspection	$P < 0.05$
2q33 karyoband	2	2	0
5q33 karyoband	8	3	2
6p21 (HLA region)	25	n.d.	10
6q21 locus	8	n.d.	0
9p13-p21 haplotype	6	n.d.	1
15q11-q13 locus	4	0	0
19p13 Lod-1 region	3	n.d.	3

[a]HC healthy control; CD MIII UT biopsy of CD patient with MIII lesion not treated by gluten-free diet; CD MIII T biopsy of CD patient with MIII lesion treated by gluten-free diet; CD MII T biopsy of CD patient with MII lesion treated by gluten-free diet; CD MI T biopsy of CD patient with MI lesion treated by gluten-free diet; CD M0 T biopsy of CD patient with M0 lesion treated by gluten-free diet; RCD MIII T biopsy of refractory CD patient with MIII lesion treated but unresponsive to gluten-free diet.
n.d. not determined.

free diet as they go through the successive stages into remission, or become refractory to the diet (Fig. 1). Molecular pathways may be identified, apart from those affected in patients, through RNAi knockouts of candidate genes in selected cell lines, or by data-mining the numerous microarray experiments that have been performed worldwide (Fig. 3).

Can identification of the perturbed molecular pathways lead us to the causative mutated gene(s)? Yes, because of the nature of the anticipated mutations, a major

role is to be expected from gene expression studies. For this we have to combine the data on differential gene expression and pathway knowledge with the positional information from the genetic studies. This can in part be achieved by integrating data from genomewide linkage analyses and microarray expression studies. Alternatively, qRT-PCR can be applied on candidate genes that have been selected through linkage and subsequent association analysis. Moreover, genes that do not show differential expression themselves might be located in relevant genetic intervals and be constituents of disturbed molecular pathways, and therefore make excellent candidate genes. We have integrated both these genetic and expression data sets in TEAM (Fig. 4), a database developed in-house with a viewing interface (Franke et al 2004).

Looking towards the future

The current status on the genetics of CD reveals a small number of interesting regions that need further investigation, including the *CELIAC2* and *CELIAC4* loci. Since these regions are rather large, verifying an allele's contribution to the disease will be a daunting task, as has become clear from the *CELIAC4* locus (Fig. 5). A systematic approach to finding all sequence variants is impractical for small research laboratories. However, much work is currently being done in the field of functional genomics. Genetic studies should try to maximize their potential by incorporating data from protein interaction networks, gene expression, binding sites for transcription factors, large-scale RNA interference studies, and gene annotation to prioritize candidate genes. Moreover, for us to take full advantage of gene expression profiles, a comprehensive understanding of the complexities of gene networks is required, as well as their transcriptional regulation. However, to finally prove the involvement of a promising candidate gene, single gene studies will remain essential.

It is expected that the total genetic risk of CD can be attributed to only one — or a few — genes with large effect (such as *HLA-DQ2*), and many genes with very modest effect (such as *CTLA4*). More robust strategies such as genomewide association may be required to identify all CD susceptibility genes with a small effect. However, these strategies require rather different study designs including 100s to 1000s of samples, and SNPs that cover the entire genome and capture all haplotype blocks. It has been estimated that whole-genome association requires at least 200 000 SNPs. It is evident that such studies come at a high price, both with respect to statistical issues (the problem of multiple testing), the large amounts of DNA required, and the high costs of genotyping. Lately, there have been great improvements in high-throughput technology, lowering costs, and improved statistics. Since disease susceptibility alleles are expected to have additive and

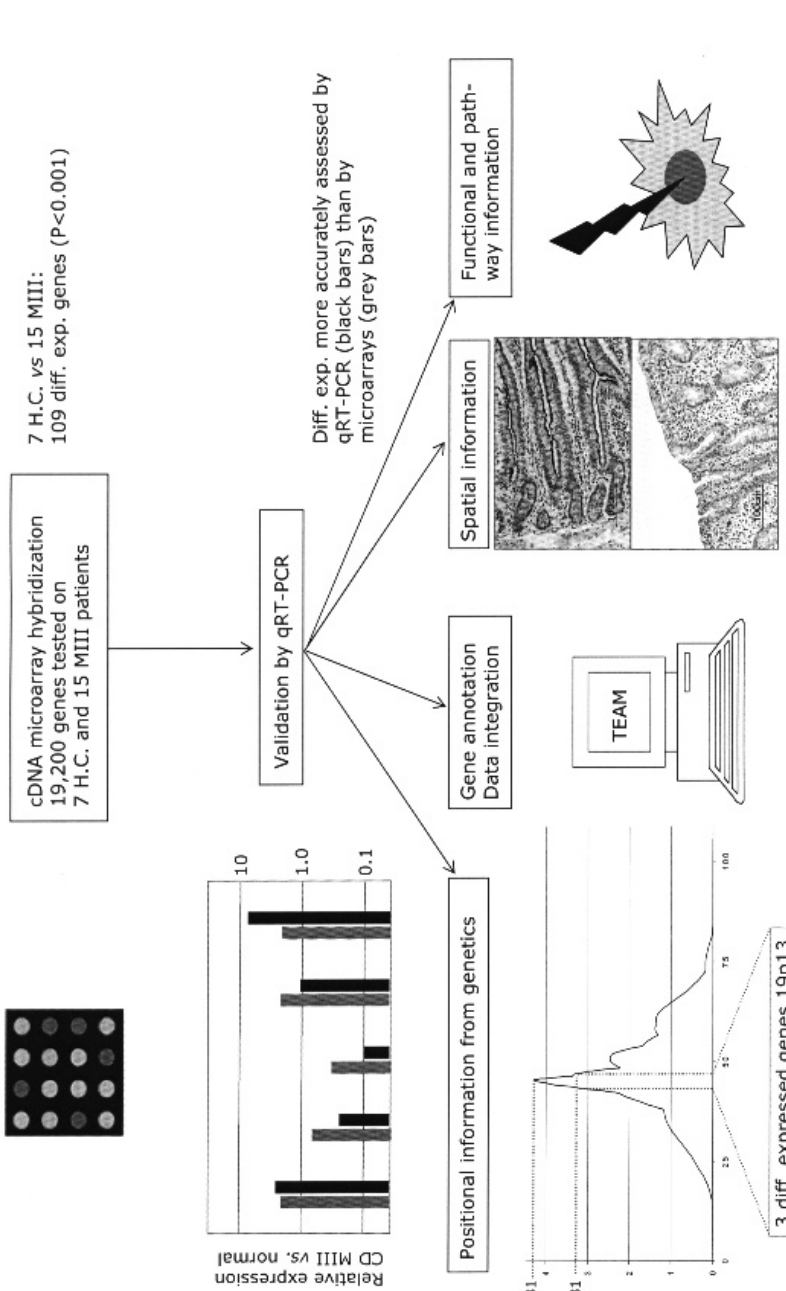


FIG. 3. Selection strategy for candidate genes based on gene expression profiling. Starting from genomewide expression analysis using microarrays, Diosdado et al (2004) identified a number of differentially expressed genes when comparing 7 healthy controls (HC) to 15 MIII biopsies from CD patients. Initial validation of differentially expressed genes should be conducted by qRT-PCR. Follow-up experiments may comprise different experimental approaches. Genetic association can be used to further investigate positional candidate genes that show differential expression and mapping to known linkage intervals. For example, the *CELIAC4* region shows three differentially expressed genes. This selection procedure is facilitated by TEAM, which permits identification of candidate genes based on the known gene functions available in public data repositories. To gain insight into the underlying disease process the behaviour of differentially expressed genes in normal, pathogenic and experimentally manipulated tissues and cells can be evaluated. For example, immunohistochemistry provides direct insight into spatial and temporal expression patterns, whereas gene knockdowns using RNAi may help identify relevant disease pathways.

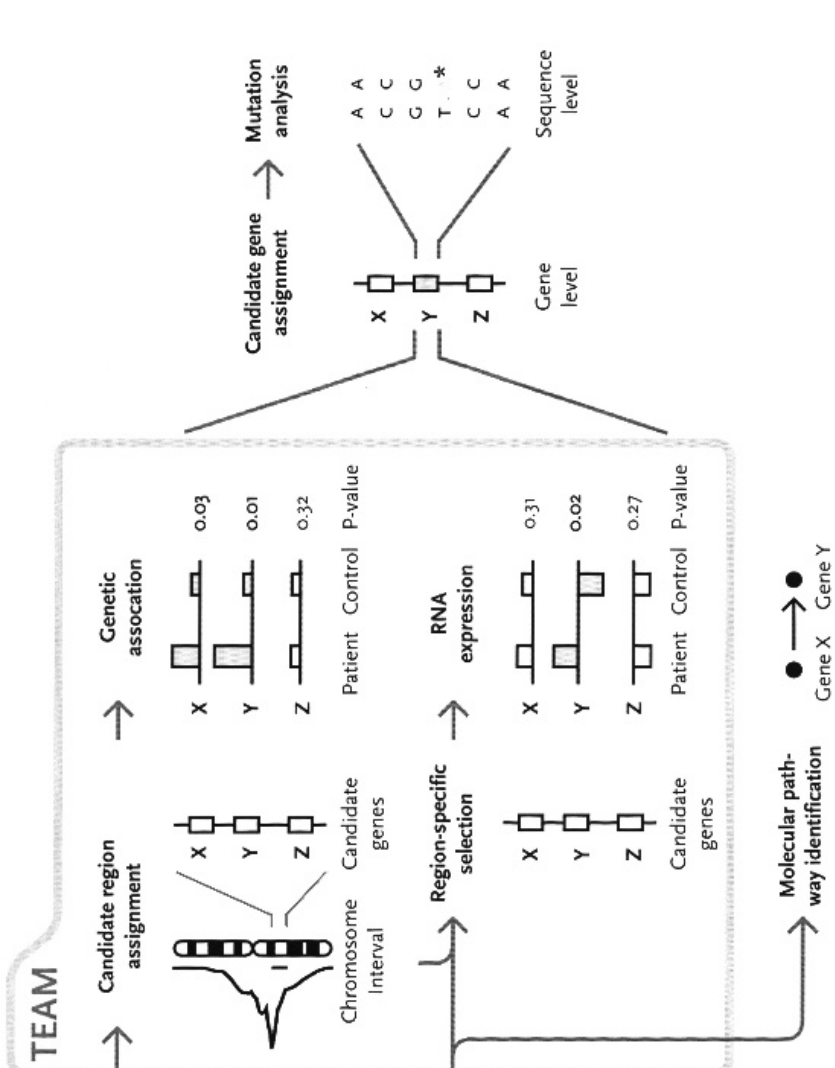


FIG. 4. TEAM is a database program that incorporates functionality for the integrated analysis of expression data, genetic data and annotation data. TEAM can be used in two ways: (1) By linking to online repositories such as Ensembl, Unigene and NCBI, we can use TEAM to help annotate genes and thereby facilitate the construction of molecular pathways; (2) by simultaneously viewing genetic linkage peaks and differentially expressed genes, we can use TEAM to facilitate selection of causative candidate genes, regardless of the availability of any functional annotation.

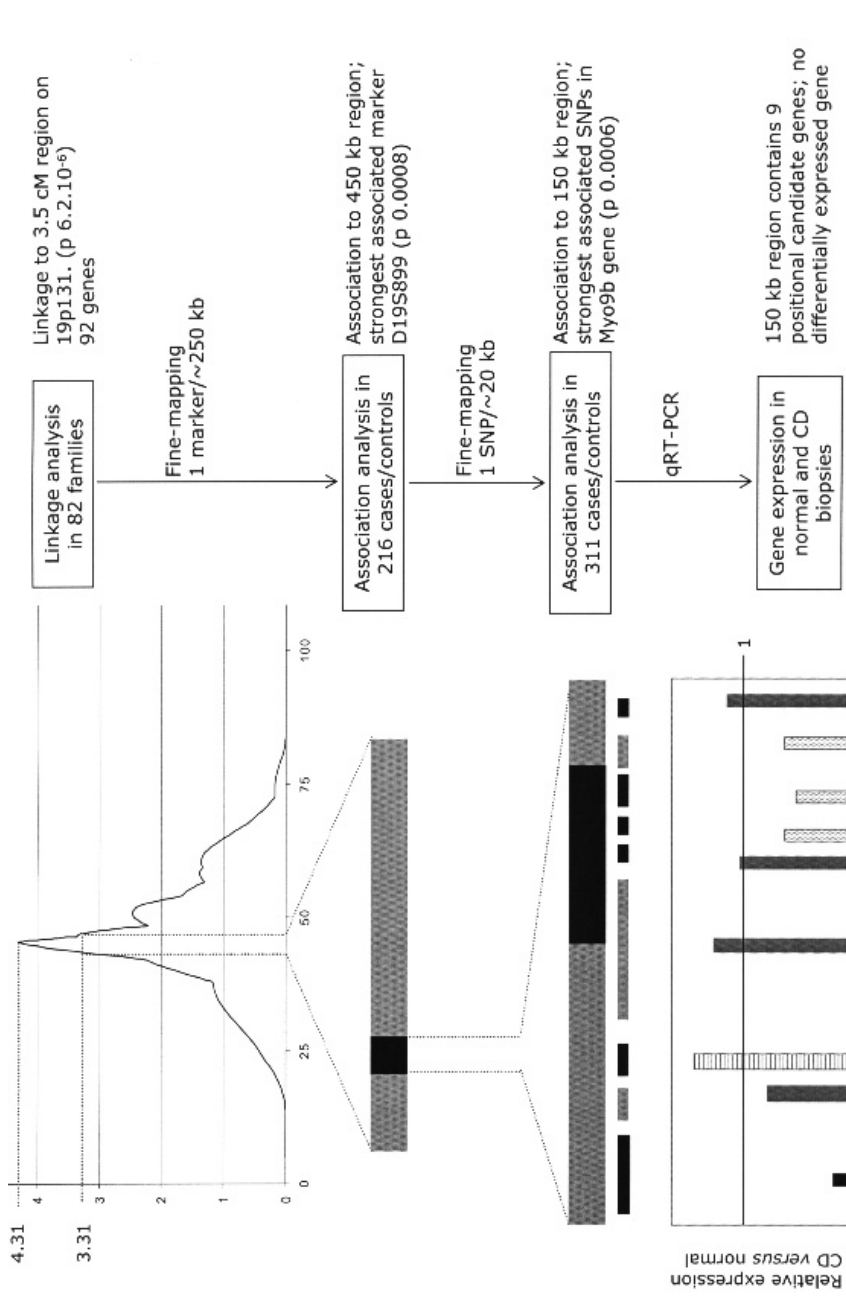


FIG. 5. The *CELIAC4* locus was mapped to chromosome 19 by linkage analysis in affected pairs of siblings. The 95% confidence interval (maximum lodscore-1) measured 3.5 cM and harboured 92 genes. Subsequent association analysis using a higher density of simple tandem repeat markers (STRPs) revealed association to a 450 kb region, which was further investigated with a high resolution of SNPs. The SNP mapping narrowed the region down to 150 kb, which encompasses nine genes. These nine genes were analysed by qRT-PCR and none of the genes showed significant differences in expression levels on comparing RNA from small intestinal biopsies of CD patients with a MIII lesion to those of healthy controls, although the levels of expression showed some variation. One gene was slightly overexpressed (indicated by horizontal stripes) and three genes were slightly down-regulated (indicated by dots). Future experiments will focus on replication studies in additional populations and the identification of a haplotype shared by CD patients form different populations. To identify additional SNPs and putative disease-causing mutations, the 150 kb region will be subjected to sequence analysis.

epistatic interactions we also need statistical tools to analyse interactions between different genes, and between genes and environmental factors.

It is anticipated that newly identified CD susceptibility genes will lead to the development of easy applicable and non-invasive molecular diagnostic tools. These tools may vastly improve the diagnosis of CD particularly as, to date, some 85% of all CD patients go undiagnosed. Better diagnostic tools are obviously only part of the solution and should be coupled with increasing awareness among general practitioners and other health care professionals. Insight into the molecular pathways involved in CD aetiology may eventually provide new targets for therapeutic intervention.

Acknowledgements

The work described in this manuscript has been made possible by grants from the Dutch Digestive Disease Foundation (WS97-44, WS00-13and WS03-06) and the Netherlands Organisation for Scientific Research (902-22-204 and 912-02-028). We thank the members of the Dutch Celiac Disease Consortium and the Complex Genetics Group at the UMC Utrecht for helpful discussions. We thank Jackie Senior for improving the manuscript.

References

Akesson E, Oturai A, Berg J et al 2002 A genome-wide screen for linkage in Nordic sib-pairs with multiple sclerosis. Genes Immun 3:279–285

Babron MC, Nilsson S, Adamovic S et al (European Genetics Cluster on Coeliac Disease) 2003 Meta and pooled analysis of European coeliac disease data. Eur J Hum Genet 11:828–834

Becker KG, Simon RM, Bailey-Wilson JE et al 1998 Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases. Proc Natl Acad Sci USA 95:9979–9984

Cox NJ, Wapelhorst B, Morrison VA et al 2001 Seven regions of the genome show evidence of linkage to type 1 diabetes in a consensus analysis of 767 multiplex families. Am J Hum Genet 69:820–830

Diosdado B, Wapenaar MC, Franke L et al 2004 A microarray screen for novel candidate genes in coeliac disease pathogenesis. Gut 59:944–951

Fahrer AM, Konigshofer Y, Kerr EM et al 2001 Attributes of gammadelta intraepithelial lymphocytes as suggested by their transcriptional profile. Proc Natl Acad Sci USA 98:10261–10266

Franke L, van Bakel H, Diosdado B, van Belzen M, Wapenaar M, Wijmenga C 2004 TEAM: a tool for the integration of expression, and linkage and association maps. Eur J Hum Genet 12:633–638

Greco L, Corazza G, Babron MC et al 1998 Genome search in celiac disease. Am J Hum Genet 62:669–675

Greco L, Babron MC, Corazza GR et al 2001 Existence of a genetic risk factor on chromosome 5q in Italian coeliac disease families. Ann Hum Genet 65:35–41

Greco L, Romino R, Coto I et al 2002 The first large population based twin study of coeliac disease. Gut 50:624–628

Haimila K, Smedberg T, Mustalahti K, Maki M, Partanen J, Holopainen P 2004 Genetic association of coeliac disease susceptibility to polymorphisms in the ICOS gene on chromosome 2q33. Genes Immun 5:85–92

Helms C, Cao L, Krueger JG et al 2003 A putative RUNX1 binding site variant between SLC9A3R1 and NAT9 is associated with susceptibility to psoriasis. Nat Genet 35:349–356

Holopainen P, Mustalahti K, Uimari P, Collin P, Maki M, Partanen J 2001 Candidate gene regions and genetic heterogeneity in gluten sensitivity. Gut 48:696–701

Holopainen P, Naluai AT, Moodie S et al (Members of the European Genetics Cluster on Coeliac Disease) 2004 Candidate gene region 2q33 in European families with coeliac disease. Tissue Antigens 63:212–222

Hugot JP, Chamaillard M, Zouali H et al 2001 Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 411:599–603

Jawaheer D, Seldin MF, Amos CI et al (North American Rheumatoid Arthritis Consortium) 2003 Screening the genome for rheumatoid arthritis susceptibility genes: a replication study and combined analysis of 512 multicase families. Arthritis Rheum 48:906–916

Juuti-Uusitalo K, Maki M, Kaukinen K et al 2004 cDNA microarray analysis of gene expression in coeliac disease jejunal biopsy samples. J Autoimmun 22:249–265

Keuning JJ, Pena AS, van Leeuwen A, van Hooff JP, van Rood JJ 1976 HLA-DW3 associated with coeliac disease. Lancet 1:506–508

King AL, Moodie SJ, Fraser JS et al 2002 CTLA-4/CD28 gene region is associated with genetic susceptibility to coeliac disease in UK families. J Med Genet 39:51–54

Liu J, Juo SH, Holopainen P et al 2002 Genomewide linkage analysis of celiac disease in Finnish families. Am J Hum Genet 70:51–59

Mora B, Bonamico M, Indovina P et al 2003 CTLA-4 +49 A/G dimorphism in Italian patients with celiac disease. Hum Immunol 64:297–301

Naluai AT, Nilsson S, Gudjonsdottir AH et al 2001 Genome-wide linkage analysis of Scandinavian affected sib-pairs supports presence of susceptibility loci for celiac disease on chromosomes 5 and 11. Eur J Hum Genet 9:938–944

Popat S, Bevan S, Braegger CP et al 2002a Genome screening of coeliac disease. J Med Genet 39:328–331

Popat S, Hearle N, Wixey J et al 2002b Analysis of the CTLA4 gene in Swedish coeliac disease patients. Scand J Gastroenterol 37:28–31

Rioux JD, Daly MJ, Silverberg MS et al 2001 Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. Nat Genet 29:223–228

Ryan AW, Thornton JM, Brophy K et al 2004 A directed candidate gene analysis of coeliac disease susceptibility on chromosome 5. In abstract book of 11th ISCD, Belfast, April 2004 (abstract no. P41).

Sakai K, Shirasawa S, Ishikawa N et al 2001 Identification of susceptibility loci for autoimmune thyroid disease to 5q31-q33 and Hashimoto's thyroiditis to 8q23-q24 by multipoint affected sib-pair linkage analysis in Japanese. Hum Mol Genet 10:1379–1386

Tokuhiro S, Yamada R, Chang X et al 2003 An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. Nat Genet 35:341–348

Ueda H, Howson JM, Esposito L et al 2003 Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 423:506–511

van Belzen MJ, Meijer JWR, Sandkuijl LA et al 2003 A major non-HLA locus in celiac disease maps to chromosome 19. Gastroenterology 125:1032–1041

Van Belzen MJ, Vrolijk M, Meijer JWR et al 2004 A genomewide screen in a four-generation Dutch family with celiac disease: evidence for linkage to chromosomes 6 and 9. Am J Gastroenterol 99:461–477

van Heel DA, Fisher SA, Kirby A, Daly MJ, Rioux JD, Lewis CM; Genome Scan Meta-Analysis Group of the IBD International Genetics Consortium 2004 Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. Hum Mol Genet 13:763–770

Woolley N, Holopainen P, Ollikainen V et al 2002 A new locus for coeliac disease mapped to chromosome 15 in a population isolate. Hum Genet 111:40–45

Xu J, Meyers DA, Ober C, Blumenthal MN et al (Collaborative Study on the Genetics of Asthma) 2001 Genomewide screen and identification of gene-gene interactions for asthma-susceptibility loci in three U.S. populations: collaborative study on the genetics of asthma. Am J Hum Genet 68:1437–1446

Zhong F, McCombs CC, Olson JM et al 1996 An autosomal screen for genes that predispose to celiac disease in the western counties of Ireland. Nat Genet 14:329–333

Chapter 2

Erfelijke factoren bij coeliakie
(in Dutch with English summary)

MC Wapenaar, AJ Monsuur en C Wijmenga

Tijdschr Kindergeneeskd 2004; **72***:11-6*

Erfelijke factoren bij coeliakie

M.C. WAPENAAR, A.J. MONSUUR EN C. WIJMENGA

SAMENVATTING

Coeliakie is een glutengeïnduceerde enteropathie met een auto-immuuncomponent. Deze relatief frequente aandoening komt geclusterd voor in risicofamilies waar ze een complexe vorm van overerving vertoont. Dit is in overeenstemming met het multifactoriële karakter van coeliakie waarbij verscheidene ziektegenen zijn betrokken. De complexe overervingspatronen verhinderen de identificatie van de verantwoordelijke genen met klassieke genetische methoden. Hier presenteren wij aangepaste methoden van genetische koppeling (sibpair-analyse) en associatie (case-control-studie) en een discussie van de hiermee behaalde resultaten. Als eerste zijn we er zo in geslaagd met statistische significantie een kandidaatregio op chromosoom 19 aan te wijzen. Voor de identificatie van het ziektegen binnen deze kandidaatregio zijn aanvullende technieken vereist die afzonderlijke behandeld worden: genexpressie, genetische associatie, mutatieanalyse en bio-informatica. Genexpressiestudie toont aan dat bij de pathologie naast immuungenen ook genen zijn betrokken die een rol spelen bij de regulatie van celproliferatie en celdifferentiatie. Voor de identificatie van de ziektegenen staan wij een strategie voor die de integratie van genetica, genexpressie en bio-informatica behelst. Ook werpen we een blik vooruit op ontwikkelingen waarvan wij verwachten dat die het onderzoek naar complexe ziekten in een stroomversnelling zullen brengen. Bundeling van dit onderzoek binnen een nationaal 'Celiac Disease Consortium' moet de efficiënte vertaling van basale onderzoeksresultaten naar toepassingen op het vlak van diagnostiek, preventie, monitoring en voeding waarborgen.

SUMMARY

Celiac disease is a gluten-induced enteropathy with an autoimmune component. This relative frequent disorder appears clustered in families at-risk where it exhibits a complex pattern of inheritance. This is in agreement with the multifactorial nature of celiac disease that involves multiple disease genes. These complex patterns of inheritance preclude the identification of the genes responsible using traditional genetic means. Here we present modified methods of linkage (sibpair analysis) and association (case/control study), together with a discussion on the results achieved. In this way, we are the first to assign with statistical significance a candidate region on chromosome 19. For the identification of the disease gene within this candidate region complementary techniques are required and will be discussed: gene expression, genetic association, mutation analysis, and bioinformatics. Gene expression study shows that in addition to immune genes, genes involved in the regulation of cell proliferation and cell differentiation take part in the pathology of celiac disease. For the identification of disease genes we propose a strategy that integrates genetics, gene expression, and bioinformatics. We also take a glance at developments that we expect to accelerate the progress in the investigation of complex diseases. Incorporation of this investigation into a national 'Celiac Disease Consortium' should warrant the results from basic research to be efficiently implemented in diagnostics, prevention, monitoring, and nutrition management.

COELIAKIE IS EEN COMPLEXE ERFELIJKE AANDOENING

Coeliakie is een klassiek voorbeeld van een multifactoriële aandoening, dat wil zeggen een ziekte die bepaald wordt door zowel genetische als omgevingsfactoren. Al meer dan 50 jaar is bekend dat een dieet zonder gluten de ziektesymptomen doet verdwijnen.[1] De genetische gevoeligheid voor coeliakie blijft echter aanwezig en introductie van gluten in het dieet zal de ziekte opnieuw induceren. De populatieprevalentie van coeliakie bedraagt 0,5-1%. De ziekte erft niet over volgens de klassieke erfelijkheidswetten (dominant, recessief). Een broer of zus van een coeliakiepatiënt heeft echter wel een verhoogde kans van 5-10% om ook ziek te worden, hetgeen overeenkomt met een relatief risico van 5-20%. Tweelingenonderzoek bevestigt dat erfelijke factoren een grote rol spelen in het ontstaan van de ziekte. De concordantie voor monozygote tweelingen bedraagt 86%, tegenover

20% voor dizygote tweelingen.[2] Dit alles wijst erop dat bij coeliakie een veelvoud aan verschillende genen betrokken zijn. Helaas geeft tweelingenonderzoek geen inzicht in het aantal en de aard van de genen die een belangrijke rol spelen in het ontstaan van coeliakie. Aangezien meer genen een rol spelen zal ieder gen afzonderlijk slechts een beperkte bijdrage leveren aan het ontstaan van de ziekte. Deze combinatie van genetische factoren kan van patiënt tot patiënt variëren. Buiten de betrokkenheid van genen is voor coeliakie vastgesteld dat omgevingsfactoren van belang zijn met een prominente rol voor gluten.

Op dit moment is van slechts één gen bewezen dat het in belangrijke mate (~40%) bijdraagt aan het ontstaan van coeliakie.[3] Dit betreft het HLA-DQ-gen, gelegen op de korte arm van chromosoom 6. Bijna alle coeliakiepatiënten brengen het zogenoemde DQ2-eiwit tot expressie, en de kleine groep DQ2-negatieve coeliakiepatiënten brengt veelal het DQ8-molecuul tot expressie. Recent is aangetoond dat deze HLA-moleculen essentieel zijn voor de presentatie van gluten aan T-cellen, hetgeen direct de functionele relatie tussen HLA-DQ2/8 en coeliakie verklaart (zie ook het vorige artikel in dit themanummer).[4,10] Deze varianten worden echter ook in ~30% van de algemene bevolking gevonden. Dit impliceert dat DQ2 en DQ8 wel noodzakelijk zijn voor het ontstaan van de ziekte, maar uitsluitend in combinatie met andere – nog onbekende – genen. Wereldwijd wordt veel onderzoek verricht om deze overige genen te vinden.

HOE SPOREN WE GENETISCHE FACTOREN OP?

Van meer dan 1200 monogene aandoeningen – veroorzaakt door mutaties in één enkel gen – zijn de afgelopen 20 jaar de ziekteveroorzakende genen opgespoord. Dit werd veelal mogelijk gemaakt door koppelingsonderzoek te verrichten binnen één grote of een aantal kleinere families waar de ziekte keurig volgens de wetten van Mendel als recessieve of dominante aandoening overerft. De locatie van een onbekend ziektegen wordt bepaald door de gezamenlijke overerving van de ziekte met een DNA-polymorfisme waarvan de chromosomale locatie wél bekend is.

Bij complexe aandoeningen zoals coeliakie is het overervingpatroon veelal niet duidelijk omdat binnen één familie meer genen een rol spelen. Hierdoor is traditioneel koppelingsonderzoek vrijwel onmogelijk.[5] Daarnaast zijn weinig uitgebreide families met coeliakie bekend. Twee veelgebruikte strategieën voor de analyse van complexe ziekten worden hierna besproken.

Associatiestudies algemeen

Associatiestudies zijn een gevestigde methode om te onderzoeken of genen met een bekende of vermeende functie een rol spelen in een ziekteproces. De achtergrond van een klassieke associatiestudie is gebaseerd op de aanwezigheid van een functionele variant in een gen die aanleiding geeft tot verschillende vormen van het genproduct: het eiwit. De hypothese die wordt getoetst is of één van de genvarianten bijdraagt aan het ontstaan van ziekte. In een associatiestudie wordt statistisch getest of die bepaalde DNA-variant significant vaker voorkomt bij een groep patiënten in vergelijking met een groep gezonde controles. De populatiefrequentie van de DNA-variant en het te verwachten verschil in distributiefrequentie tussen de groepen is bepalend voor de noodzakelijke grootte van de patiënt- en controlecohorten.

Associatiestudies in coeliakie

Kandidaatgenen worden gevonden door te kijken naar de processen die bij een ziekte een rol spelen. Voor coeliakie is dit lastig, aangezien alle bij de ziekte betrokken processen nog niet duidelijk in beeld zijn gebracht. Alleen genen die een rol spelen in de immuunrespons zijn voor de hand liggende functionele kandidaatgenen. Tot op heden zijn relatief weinig genen getest in associatiestudies (tabel I).

Het gen dat de meeste aandacht trekt is cytotoxisch-T-lymfocyt-antigen-4 (CTLA4). CTLA4 is betrokken bij de negatieve regulatie van de T-celrespons. Associatiestudies naar CTLA4 laten de betrokkenheid van dit gen zien in diverse auto-immuunziekten, waaronder type-1-diabetes. Het belang van dit gen ligt dan ook waarschijnlijk meer bij auto-immuniteit in het algemeen.[6] Dit gen is niet alleen op grond van bovenvermelde functie een geschikt kandidaat-

Tabel I: Kandidaatgenen getest voor associatie met coeliakie

gen	*populatie*	*p-waarde*[a]
CTLA4	Finnen	0,0001
	Zweden/Noren	0,007
	Britten	0,007
	Noord-Europeanen	0,039
	Zweden	0,02
	Italianen	0,03
TNF	Ieren	<0,001[b]
	Ieren	<0,000001[b]
	Finnen	<0,0001[b]
tTG	Nederlanders	n.s.
IL12B	Scandinaviërs	n.s.
	Italianen	n.s.
MICA	Spanjaarden	<0,0001[b]
	Spanjaarden	0,02

[a] n.s. = niet significant.
[b] Associatie komt waarschijnlijk door het DQ2-haplotype.

Tabel II: Kandidaatgebieden voor coeliakiegenen. Resultaten van koppelingsonderzoeken in diverse Europese populaties

chromosoom-gebied[a]	*oorspronkelijke populatie*	*mogelijke koppeling*	*suggestieve koppeling*	*significante koppeling*	*replicatie populaties*
2q33	Finnen		x		Finnen Zweden/Noren Noord-Europeanen
4p15	Finnen		x		Finnen Noord-Europeanen
5q33	Italianen	x			Italianen Scandinaviërs Finnen
6p23	Ieren		x		Britten
9p21	Scandinaviërs	x			Finnen Nederlanders
11p11	Iers		x		Britten (2x)
15q12	Finnen			x	
19p13	Nederlanders			x	

*Gebieden met een hoge significantie of gebieden die herhaald zijn gevonden.

gen, maar ook op basis van de chromosomale locatie, aangezien de regio waar dit gen ligt (2q) al vaker gevonden is in koppelingsonderzoek (tabel II). Een aantal studies – uitgevoerd in diverse bevolkingsgroepen – laat inderdaad associatie zien tussen CTLA4 en coeliakie. De DNA-varianten waarmee associatie wordt gevonden verschillen tussen de populaties. Recent onderzoek laat echter zien dat de DNA-variant in CTLA4 die tot nu toe het meest is bekeken (+49A/G) wellicht niet het sterkst is geassocieerd.[6]

Voor twee andere kandidaatgenen (IL12B en tTG) kon geen associatie worden aangetoond. Dit betekent niet dat deze genen geen rol spelen bij de pathologie van coeliakie, maar dat een causale rol vooralsnog is uitgesloten. Voor andere genen (TNF en MICA) is het moeilijk om associatie vast te stellen, doordat deze genen in de buurt van het HLA-DQ-gen liggen en niet goed te onderscheiden is wat de precieze invloed van HLA is.

Deze resultaten tonen aan dat het lastig is om een goede voorspelling te doen over de geschiktheid van een kandidaatgen zolang we geen inzicht hebben in hoe coeliakie precies ontstaat en welke cellulaire processen daarbij betrokken zijn.

Koppelingsonderzoek in broer/zus paren ('sibpairs')

Hoewel traditioneel koppelingsonderzoek met grote families niet mogelijk is bij coeliakie, kunnen we een variatie op deze methode wel toepassen op kleine families, uitsluitend bestaande uit aangedane broer-zusterparen ('sibpairs'). Op grond van het herhalingsrisico van coeliakie zal 5-10% van alle patiënten een broer of zus hebben die ook de ziekte heeft. Gemiddeld delen broer en zus 50% van hun erfelijk materiaal. Als beiden dezelfde ziekte hebben, zullen ze genetisch identiek zijn voor dát deel van het chromosoom waarop het verantwoordelijke ziektegen ligt. Het DNA van een groot aantal sibpairs (100-200) wordt getest met ~300 DNA-polymorfismen. Voor het merendeel van die 300 DNA-polymorfismen zullen de sibpairs 50% genetisch identiek zijn, behalve voor dat DNA-polymorfisme dat dicht in de buurt ligt van het ziektegen. Daar zullen de sibpairs dan 100% identiek zijn (fig. 1). Maar omdat er meer ziektegenen zijn, zal altijd maar een beperkt aantal families dezelfde koppeling tonen. Om redenen van statistische significantie is het dan ook noodzakelijk voor deze studies honderden families te gebruiken. Dit koppelingsonderzoek leidt tot het vinden van chromosomale gebieden van beperkte omvang waarin gemiddeld 50-500 genen liggen die alle op basis van hun locatie in het genoom als kandidaatgenen beschouwd moeten worden.

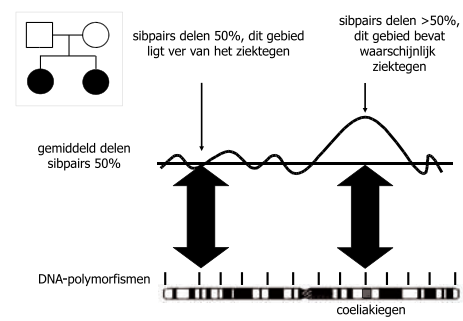


Fig. 1. Schematisch overzicht van genetisch koppelingsonderzoek. Door DNA-polymorfismen verspreid over alle chromosomen te testen wordt gekeken of sibpairs bepaalde gebieden meer dan gemiddeld 50% delen. Deze gebieden bevatten waarschijnlijk genen die primair betrokken zijn bij coeliakie.

WELKE GEBIEDEN ZIJN ER UIT DE KOPPELINGSSTUDIES GEKOMEN?

De eerste studie werd uitgevoerd door Zhong in 1996 in 45 sibpairs afkomstig van 15 Ierse families. Sinds 1996 zijn er 11 totale-genoomstudies uitgevoerd. Zeven hiervan betroffen sibpair-analyses, waaronder de Nederlandse studie van Van Belzen e.a, die in hun artikel ook een overzicht geven van de belangrijkste gevonden gebieden.[7] Daarnaast zijn enkele studies met uitgebreidere families verricht. Aan de hand van de gevonden resultaten hebben andere groepen zich specifiek op die gebieden met koppeling gericht in de hoop die uitkomsten te repliceren om zo het statistische bewijs te versterken. Behalve dat iedere studie koppeling vond in het HLA-gebied, zijn ook een aantal andere gebieden herhaald gevonden (tabel II). Op chromosoom 2 (2q33) is door Holopainen in 102 Finse families (totaal 140 sibpairs) een gebied gevonden, hetgeen is bevestigd in drie andere studies.[8] Geen van de vier studies afzonderlijk leverde significante resultaten op, maar het feit dat het herhaald is gevonden maakt deze regio toch aantrekkelijk, temeer daar dit gebied het CTLA4-gen bevat. Veel groepen hebben hun aandacht gericht op chromosoom 5 (5q31-33). Dit gebied wordt herhaaldelijk bevestigd in genetische studies, maar met een lage statische betrouwbaarheid. Dit gebied is interessant omdat het een groot aantal kandidaatgenen bevat, zoals cytokinen.

Recent hebben wij een genetische studie afgerond in 84 families (101 sibpairs) waarbij strenge diagnostische selectiecriteria zijn aangehouden en alleen patiënten met een bewezen (sub)totale vlokatrofie van de dunne darm (Marsh-III-classificatie) zijn meegenomen.[7] Dit is de eerste studie die een significant resultaat laat zien en waarbij koppeling is gevonden met chromosoom 19, dit in een gebied dat ongeveer 100 genen bevat. Daarnaast vonden Van Belzen e.a. ook sterke aanwijzingen voor een kandidaatgen op chromosoom 6, niet gerelateerd aan het HLA-gebied. Dit gebied is mogelijk ook betrokken bij type-1-diabetes. Verder onderzoek zal moeten uitwijzen of hier een gen ligt dat betrokken is bij auto-immuniteit.

Na aanwijzing van een aantal interessante chromosomale gebieden (tabel II) volgt de speurtocht naar de daarin gelegen coeliakiegenen. Een aantal verschillende technieken is voorhanden die behulpzaam kunnen zijn bij dit identificatieproces. Veelal wordt in wisselende combinaties gebruik gemaakt van de volgende methoden: genexpressiestudies, genetische associatie, bio-informatica, mutatieanalyse en diermodellen. Welke combinatie wordt gekozen wordt in ieder specifiek geval bepaald door de middelen die voorhanden zijn en die gedurende de loop van het onderzoek kunnen variëren (fig. 2).

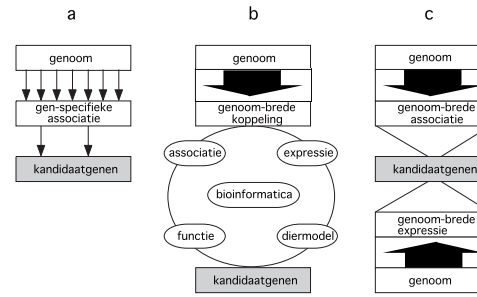


Fig. 2. Schematisch overzicht van drie strategieën voor de selectie van kandidaatgenen voor complexe genetische ziekten zoals coeliakie: a) de a priori selectie van genen die getest worden door middel van associatie is arbitrair en beperkend maar een veelgebruikte benadering; b) de door ons verkozen benadering gaat uit van een koppelingsstudie zoals een sibpair-analyse waarbij alle genen in het genoom worden betrokken; vanuit die groepen genen die overerven met de ziekte worden kandidaatgenen geselecteerd waarbij gebruik gemaakt wordt van een combinatie van de aangegeven methodieken; c) de geïntegreerde benadering van genoombrede associatie (high density SNP's) en expressiestudies (microarrays) zijn momenteel technologisch nog een tour de force maar binnen afzienbare tijd binnen bereik.

GENEXPRESSIESTUDIE

Mutaties in een ziektegen leiden doorgaans tot veranderingen in de aanmaak, stabiliteit, vorm of functie van het genproduct, het eiwit, via de veranderingen in de overgeschreven boodschapper-RNA-moleculen. Het meten van deze kwantitatieve en kwalitatieve veranderingen in het RNA en/of eiwit in het aangedane weefsel geeft in principe direct toegang tot het verantwoordelijke ziektegen. Mutaties in ziektegenen brengen echter ook verstoringen teweeg in de moleculaire routes waarbinnen zij functioneren, met als resultaat ontregelde cellulaire processen die doorwerken op orgaan- en systemisch niveau. Zo kan een mutatie in één ziektegen leiden tot een kettingreactie van passieve kwantitatieve veranderingen in de expressie van de andere betrokken genen. Vergelijken we dan, in het geval van coeliakie, genexpressieveranderingen in het duodenum van een patiënt ten opzichte van een gezonde controle, dan vinden we zoveel verschillen dat de specifieke veranderingen van het verantwoordelijke ziektegen obscuur blijven. Echter, als we deze

uitkomst combineren met de resultaten van het genetisch koppelingsonderzoek dan ontstaat een totaal ander beeld. Vergelijken we de groep genen geselecteerd op basis van expressieveranderingen met de groep geselecteerd op grond van genetische (positionele) informatie dan levert dit een beduidend kleinere deelverzameling van kandidaatgenen op die voldoen aan beide criteria.

Momenteel bestuderen wij zo de genen in de door ons geïdentificeerde kandidaatregio op 19p13 voor veranderingen in RNA-expressie. Hierbij maken we gebruik van de 'real-time reverse transcription polymerase chain reaction' (RT-PCR)-techniek, een uiterst gevoelige detectiemethode die ons in staat stelt met de minieme hoeveelheden RNA te werken die we uit de duodenumbiopten kunnen extraheren. In samenhang met de hierna te bespreken methoden trachten we zo het coeliakierisicogen in 19p13 te identificeren.

Voor de RNA-expressiestudies beperken we ons echter niet tot de 19p13-kandidaatregio. In samenwerking met dr. Mulder (VUmc) en dr. Mearin (LUMC) is een grote collectie duodenumbiopten opgebouwd, niet alleen van patiënten en controles, maar ook van de verschillende ziektestadia in respons op een glutenvrij dieet, en tevens van de zeldzamere refractaire patiënten die ongevoelig zijn voor dit dieet. Met deze collectie hebben we zo toegang tot de moleculaire veranderingen die aan de basis staan van respectievelijk de lymfocytose, de crypthyperplasie, de villeuze atrofie en dieetrespons. Omdat bij al deze veranderingen mogelijk honderden genen betrokken zijn, hebben we in samenwerking met dr. Holstege (UMCU) de microarray-technologie geïmplementeerd, die het mogelijk maakt de verandering in expressie van tienduizenden genen simultaan te detecteren. Op deze wijze is het mogelijk om op het niveau van het hele genoom alle veranderingen te meten die gepaard gaan met de pathologische transformatie van de duodenale mucosa. Onze eerste resultaten wijzen erop dat die gekenmerkt wordt door een verhoogde celproliferatie bij een achterblijvende celdifferentiatie.[9]

In samenwerking met dr. Koning (LUMC) worden in biopten, parallel aan de RNA-kwantificeringen, ook eiwitmetingen verricht. Immunohistologie kan verder inzicht verschaffen of de veranderingen in genexpressie hun oorsprong vinden in de activiteit van de individuele cellen zelf, of dat er sprake is van toename van het aantal cellen en/of veranderingen in weefseldistributie. 'Laser-scanning microscopy' maakt het mogelijk microdissectie uit te voeren op biopten en vervolgens de RNA-expressie te kwantificeren. Deze geïntegreerde benadering zal uiteindelijk leiden tot het ontraadselen van de moleculaire pathogenese en, mits gekoppeld aan de genetische resultaten, de identificatie van de ziektegenen en de betrokken cellulaire routes bespoedigen.

GENETISCHE ASSOCIATIE

De a priori keuze van de te testen kandidaatgenen in de klassieke genetische associatiestudie is arbitrair en beperkend. Echter, met de resultaten van het koppelingsonderzoek richten we onze aandacht uitsluitend op die gebieden in het genoom waar aanwijzingen voor genetische factoren zijn. Dit heeft als voordeel dat we die kandiaatregio's kunnen bemonsteren met een hoge dichtheid aan DNA-polymorfismen. Door de inspanningen van het Humane Genoom Project zijn miljoenen DNA-polymorfismen beschikbaar gekomen in de vorm van enkele-basevariaties, de zogenoemde 'single nucleotide polymorphisms' (SNP's). In de kandidaatregio op 19p13 passen we deze associatieanalyse met SNP's toe om dit interval verder te verkleinen tot een gebied ter grootte van enkele kandidaatgenen.

MUTATIEANALYSE

Door het bepalen van de basenvolgorde in het DNA van de genen, de sequentieanalyse, kunnen we die varianten detecteren die verantwoordelijk zijn voor het disfunctioneren van het ziektegen. Duidelijke voorbeelden zijn die mutaties die leiden tot de vroegtijdige terminatie van de peptidenketen, of de inbouw van verkeerde aminozuren. Mutaties in de promotor van het gen kunnen verminderde expressie tot gevolg hebben. Direct na de transcriptie wordt tijdens het proces van 'RNA-splicing' die delen van het RNA met elkaar verbonden die coderen voor het eiwit (exonen), terwijl de niet-coderende delen (intronen) verwijderd worden. Mutaties kunnen dit proces verstoren. Mutatieanalyse speelt een cruciale rol bij de identificatie van ziektegenen maar leent zich niet voor screening van grote genomische regio's voor kandidaatgenen.

BIO-INFORMATICA

Uit het voorgaande is duidelijk dat verschillende methodieken bijdragen aan het opsporen van kandidaatgenen, maar dat de kracht vooral gelegen is in de synergie van deze benaderingen. Integratie veroorzaakt een enorme datastroom die met behulp van computers en bio-informatica niet alleen verwerkt moet worden, maar vooral inzichtelijk moet blijven. De functionele annotatie van genen neemt in hoog tempo toe en wordt gedeponeerd in publiekelijk toe-

gankelijke databankbestanden. Ook deze informatie is essentieel bij de selectie van kandidaatgenen. Binnen onze groep is een programma ontwikkeld (TEAM) dat deze informatiestroom inzichtelijk maakt en behulpzaam is bij het maken van keuzes in de onderzoeksstrategie.

HIGH-THROUGHPUT TECHNIEKEN

De gelimiteerde resolutie van koppelingsonderzoek is het gevolg van het beperkt voorhanden zijn van families met aangedane 'sibpairs' en de onderliggende genetische structuur. Associatiestudies hebben dit nadeel niet, maar kennen geen hypothesevrije benadering door de arbitraire selectie van kandidaatgenen. Ideaal is een associatiestudie met een hoge dichtheid aan DNA-polymorfismen (SNP's) verspreid over het hele genoom. Dit behelst het bepalen van de genotypen van honderden patiënten en controles met honderdduizenden SNP's. Beperkingen zijn momenteel de voorhanden zijnde technologie en het kostenaspect. De ontwikkelingen zijn zodanig dat dit binnen afzienbare tijd tot de mogelijkheden behoort. Gecombineerd met de genoomwijde expressiestudies met behulp van microarrays komen high-throughput screening en identificatie van risicogenen in complexe ziekten zoals coeliakie binnen bereik (fig. 2).

ONTWIKKELINGEN EN VOORUITZICHTEN

Gezien de snelheid van de technologische ontwikkelingen en de groeiende inzichten uit het humane genoomonderzoek is het huidige tijdvak rijp voor het ontrafelen van de genetica en moleculaire pathologie van coeliakie. Voorwaarde hierbij is wel dat we over grote aantallen patiënten en controlemonsters beschikken in de vorm van DNA uit bloed en RNA uit duodenumbiopten. De lopende samenwerking tussen het LUMC, VUmc en UMCU, maar ook het contact met de patiënten en hun familieleden via hun belangenverenigingen is hiervoor onontbeerlijk. Doelstelling is om bij de identificatie van risicogenen en/of biomarkers voor de pathologie deze aan te wenden voor verbeterde diagnostiek en prognose van het ziekteverloop, en ter verdere ondersteuning tijdens de behandeling.

Inzicht in de moleculaire pathologie opent ook de mogelijkheid voor gerichte farmacologische interventie. De relatief hoge incidentie van coeliakie, gekoppeld aan het aanzienlijke aantal onopgemerkte patiënten maakt dat bij beschikbaarheid van geschikte tests een bevolkingsonderzoek zeker overwogen moet worden. Bestudering en behandeling van coeliakie bestrijkt het hele spectrum van genetische aanleg tot voeding. Vanuit die overtuiging is een nationaal multidisciplinair consortium van onderzoekers gevormd met vertegenwoordigers van het UMCU, VUmc, LUMC, RIKILT (Wageningen) en RIVM (Bilthoven) om kennis te bundelen en ervaringen uit te wisselen.

DANKBETUIGINGEN

De auteurs wensen de volgende medewerkers van de Complexe Genetica Groep te bedanken: Martine van Belzen, Begoña Diosdado en Lude Franke. Het onderzoek van ons team wordt mede mogelijk gemaakt door financiële bijdragen van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) en de Maag Lever Darm Stichting (MLDS). Verder zijn wij erkentelijk voor de belangenloze medewerking van de patiënten en hun familieleden zonder wie dit onderzoek niet mogelijk is.

Martin C. Wapenaar, mw. Alienke J. Monsuur en mw.prof.dr. Cisca Wijmenga, Complexe Genetica Groep, Divisie Biomedische Genetica, Universitair Medisch Centrum Utrecht.
Correspondentieadres: Prof.dr. Cisca Wijmenga, Complexe Genetica Groep S2.112, Divisie Biomedische Genetica, UMC Utrecht, Postbus 80030, 3508 TA Utrecht, e-mail: t.n.wijmenga@med.uu.nl.

LITERATUUR

1 Dicke WK. Coeliakie. Een onderzoek naar de nadelige invloed van sommige graansoorten op de lijder aan coeliakie. Proefschrift. Rijksuniversiteit Utrecht, 1950.
2 Greco L, Romino R, Coto I, et al. The first large population based twin study of coeliac disease. Gut 2002;50(5):624-8.
3 Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. Nat Rev Immunol 2002;2(9):647-55.
4 Wal Y van de, Kooy Y, Veelen P van, et al. Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. J Immunol 1998;161(4):1585-8.
5 Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet 2003;33(Suppl):228-37.
6 Ueda H, Howson JM, Esposito L, et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 2003;423(6939):506-7.
7 Belzen MJ van, Meijer JW, Sandkuijl LA, et al. A major non-HLA locus in celiac disease maps to chromosome 19. Gastroenterology 2003;125(4):1032-41.
8 Holopainen PM, Partanen JA. Technical note: linkage disequilibrium and disease-associated CTLA4 gene polymorphisms. J Immunol 2001;167(5):2457-8.
9 Diosdado B, Wapenaar MC, Franke L, et al. A microarray screen for novel candidate genes in coeliac disease pathogenesis. Gut 2003;in press.
10 Koning F, De immunologie van coeliakie. Ned Tijdschr Kindergkd 72;2004:0-00.

SECTION II

HIGH-THROUGHPUT GENE EXPRESSION ANALYSIS

Chapter 3

Gene expression mosaicism precedes duodenal tissue patchiness in coeliac disease

MC Wapenaar, M Hadithi, M van Kollenburg,
GA Meijer, C Wijmenga, and CJJ Mulder

(Submitted for publication)

ABSTRACT

Background & Aims: Diagnosis, treatment monitoring, and gene expression research in coeliac disease (CD) rely on accurate Marsh histological classifications. Patchy lesions and molecular heterogeneity may interfere with these applications. Our aim was to assess the impact of these variables on gene expression. ***Methods***: In total, 15 patients, including cases of *de novo* CD (n=4), treated CD (n=4), refractory CD (n=3), dermatitis herpetiformis (n=4), and 8 non-CD controls participated in the study. Six pairs of biopsies were sampled from each individual and from separate duodenal folds, and examined for histology and gene expression. Expression of *IFNG* and *TM4SF4*, as measured by qRT-PCR, were used a markers to reflect mucosal inflammation and differentiation, respectively. ***Results***: Patients with more severe lesions generally showed increased *IFNG* and decreased *TM4SF4* expression. Patchy lesions were mainly observed in refractory and *de novo* CD patients. Remarkably, patients without tissue patchiness also showed considerable and comparable expression heterogeneity for which about one/third could be attributed to technical variations. Despite this, we found strong correlations between the Marsh classification and gene expression in individual biopsies, irrespective of the patients they originated from. Strong correlation (R^2 = 0.89) was observed between the ratio of *IFNG* and *TM4SF4* expression and the weighted averaged Marsh classification. ***Conclusions***: Expression heterogeneity is caused by 1) inter-individual variability, 2) mucosal heterogeneity, and 3) technical variations. Expression heterogeneity appears to precede tissue patchiness. Despite these variables, Marsh lesions and *IFNG*/*TM4SF4* expression were strongly correlated, independent of disease diagnosis. We suggest that expression studies should be performed with 10-20 samples per Marsh class from separate individuals to cope with distorting variables.

INTRODUCTION

Coeliac disease (CD) is a chronic inflammatory condition of the small intestine caused by an autoimmune response to dietary gluten in genetically susceptible individuals[1]. CD is a complex genetic disorder, and with a frequency of ~1% in Western populations, it is among the more common multifactorial diseases[2]. Patients display a range of mucosal lesions that can be ranked according to the modified Marsh classification and progressively include lymphocyte infiltration (MI), crypt hyperplasia (MII), and villous atrophy (MIII)[3]. The latter ranges from partial (MIIIa), and subtotal (MIIIb), to total atrophy (MIIIc)[4]. Patients may go into complete remission (M0) on a gluten-free diet (GFD) and then appear similar to healthy controls. Identical Marsh lesions may be observed in patients that are unresponsive to a GFD (refractory CD, RCD), and in some patients with the gluten-sensitive skin condition, dermatitis herpetiformis (DH). Current diagnostic

tools include tTG and EMA serology[5], and HLA-DQ-typing, but endoscopy with duodenal biopsy examination is still considered the 'gold standard'[6]. In addition to histological restructuring of the mucosa, molecular events may also reflect the progression of CD lesions. Previously, we have shown that increased expression of the gene for the proinflammatory cytokine interferon-gamma (*IFNG*) coincided with more severe lesions[7]. Likewise, we have identified the brush-border gene *TM4SF4* (transmembrane 4 L six family member 4) as a terminal differentiation marker that decreased in expression when the disease progressed towards higher Marsh categories[8]. Recently, we used high-throughput gene expression profiling methods, like microarrays and qRT-PCR, to further dissect the inflammatory and differentiation pathways in CD pathogenesis (Diosdado, submitted; Wapenaar, submitted). Both the initial diagnosis and monitoring of treatment results in the clinic, as well as gene expression studies in basic research, rely on accurate Marsh classification of endoscopically collected biopsy samples. However, a patchy distribution of mucosal lesions may interfere with a correct Marsh classification and have been reported repeatedly[9-11]. Here we present a detailed study on mucosal heterogenity, both on the histological and the molecular level, by comparing Marsh typing with *IFNG* and *TM4SF4* expression patterns. We discerned various types of variation and will discuss their impact on expression studies. We provide general guidelines on how to deal with tissue patchiness and expression heterogeneity that may also prove useful in other areas than research and diagnostics of CD.

PATIENTS/MATERIALS AND METHODS

Patients and biopsy sampling

The 23 individuals that participated in this study were enrolled between 2004 and 2005 by the Department of Gastroenterology and Hepatology at the VU University Medical Center, Amsterdam. Patients included were: newly diagnosed CD (n=4), CD in remission on a GFD (n=4), RCD on a GFD (n=3), DH (n=4) of which one followed a GFD, and non-CD controls with gastroesophageal reflux disease (n=8). See table 1 for further details. All patients underwent upper gastrointestinal endoscopy under conscious sedation (using fentanyl and midazolam). After introducing the endoscope as deep as the pars ascendens of the duodenum, six pairs of biopsy samples were taken from different duodenal folds starting with the deepest one in the proximal part. The distance between samples within a single biopsy pair was approximately 5 - 10 mm. In patients with a loss of folds, biopsies were taken from six levels with approximately 2 cm distance. From each pair of biopsies harvested, one sample was set aside for histological examination, and the other sample was preserved in RNA*later* (Ambion, Austin, TX) at –20°C for subsequent RNA isolation. This study was approved by the Medical Ethics Committee of the UMC Utrecht and all the participants gave their informed consent.

Histology

Characteristics of the Marsh categories

Marsh stages	Values	Characteristics
M0	Value 1	normal/complete remission
MI	Value 2	lymphocytosis
MII	Value 3	additional crypt hyperplasia
MIIIa	Value 4	additional partial villous atrophy
MIIIb	Value 5	additional subtotal villous atrophy
MIIIc	Value 6	Additional total villous atrophy

Gene expression analysis

Total RNA isolation, cDNA synthesis, and qRT-PCR were performed essentially as described previously[7]. In brief, RNA was isolated from a single biopsy in 500 ul TRIzol (Invitrogen, Carlsbad, CA) by homogenization with a Mini-BeadBeater (BioSpec Products Inc., Bartlesville, OK). RNA quality and quantity was checked on a 2100 Bioanalyzer (Agilent, Palo Alto, CA). All reagents and equipment for qRT-PCR were purchased from (Applied Biosystems, Foster City, CA). Synthesis of cDNA was carried out with the High Capacity Archive Kit. *IFNG* transcripts were detected with a FAM-labeled Taqman probe specific for exon 2 using the Assay-on-Demand Hs00174143m1. *GUSB* transcripts were detected with the VIC-labeled PDAR 4362320E and served as an endogenous control. Reactions were performed using the Universal PCR Master Mix. Transcripts of *TM4SF4* were quantified using the forward primer 5'- TCT TGG GCC TGA AGA ACA ATG –3' and reverse primer 5'- AGC AAA TAT CGT GGA GGT GAA CA –3'. *GUSB* served as endogenous control by using the forward primer 5'- GAA AAT ATG TGG TTG GAG AGC TCA TT – 3' and reverse primer 5'- CCG AGT GAA GAT CCC CTT TTT A – 3'. Products were detected by SyberGreen incorporation using the SyberGreen PCR Master Mix. PCR cycling and real-time data collection was performed with an SDS 7900HT. Relative expression was calculated with the delta(delta(Ct))-method[12], using the Ct-values obtained with SDS2.2 software. We used the averaged delta(Ct)-value of the six controls with homogenous M0 lesions (patients A-F) as a calibrator to calculate the normalized relative expression for all samples.

RESULTS

For this study on patchy mucosal lesions in coeliac disease, we had 23 participants. The clinical and immunological data, together with information on treatment and associated diseases, are summarized in table 1. Closely spaced pairs of biopsy samples were collected from each individual from six different folds in the distal duodenum. These paired biopsies were processed and examined separately for histology and gene expression. Progression of the mucosal lesions was indicated

Table 1. Clinical and immunological details on the 15 patients and 8 controls used in this study

Individual	Sex	Age	HLA	tTG	EMA	GFD	Diagnosis	Complica-tions
A	M	51	DQ2/-	-	-	-	NC	AITD
B	F	25	DQ2/-	-	-	-	NC	UC
C	M	60	DQ2/-	-	-	+	NC	
D	M	43	DQ2/-	-	-	-	NC	
E	M	36	DQ2/-	-	-	-	NC	
F	F	71	DQ8/-	-	-	-	NC	
G	F	42	DQ2/-	-	-	-	NC	
H	M	51	DQ2/-	-	-	-	NC	HP
I	F	72	DQ2/-	-	-	+	CD	
J	F	77	DQ2/-	-	+	+	CD	
K	M	20	DQ2/-	-	-	+	CD	IGAD, SS
L	M	43	DQ2/DQ2	-	-	+	CD	
M	M	61	DQ2/-	-	-	-	DH	
N	F	63	DQ2/-	-	+	-	DH	
O	M	41	DQ2/-	-	-	-	DH	
P	M	58	DQ2/-	-	+	+	DH	
Q	M	83	DQ2/-	+	+	-	CD	
R	F	61	DQ2/-	+	+	-	CD	
S	M	43	DQ2/-	+	+	-	CD	FA
T	F	36	DQ2/-	+	+	-	CD	
U	F	64	DQ2/-	-	-	+	RCD	AITD
V	M	66	DQ2/DQ2	-	-	+	RCD	
W	F	66	DQ2/DQ2	+	-	+	RCD	EATL

Abbreviations (in order of appearance): M, male; F, female; HLA, major histocompatibility class II DQ-type; tTG, anti-tissue transglutaminase; EMA, anti-endomysium; GFD, gluten-free diet; NC, non-CD control; CD, coeliac disease; DH, dermatitis herpetiformis; RCD, refractory coeliac disease; AITD, autoimmune thyroiditis; EATL, enteropathy-associated T-cell lymphoma; FA, Fanconi anemia; HP, Helicobacter pylori gastritis; IGAD, IgA deficiency; SS, Sjögren's syndrome; and UC, ulcerative colitis.

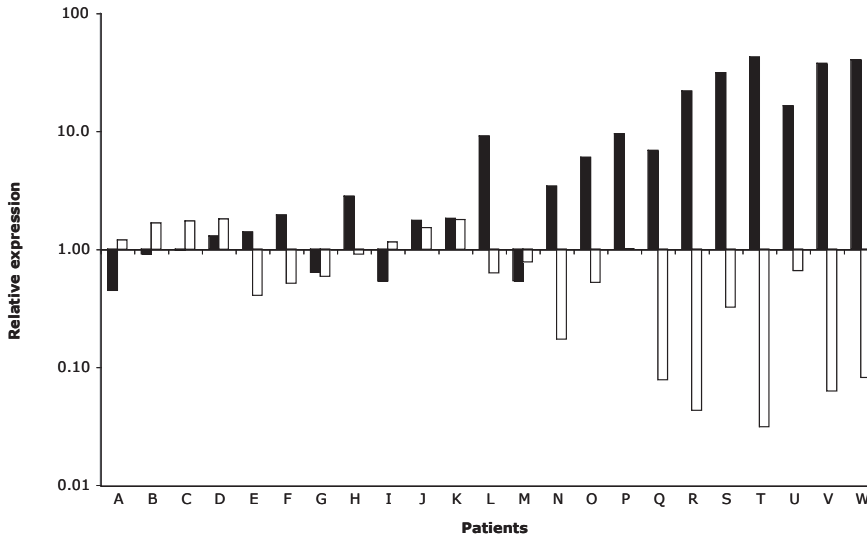


Figure 1. Relative gene expression of *IFNG* (filled bars) and *TM4SF4* (open bars) in 23 patients and controls. Individuals are indicated with capital letters A-W (see table 1 for details). Vertical bars represent the relative expression calculated from the average of six replicate biopsy samples (see table 2). Each biopsy sample was tested in triplicate. Expression values were normalized to the averaged expression of the controls A-F.

using the Marsh nomenclature, while expression of the genes *IFNG* and *TM4SF4* were used as markers for mucosal inflammation and differentiation, respectively. These results are compiled in table 2. As expected, the progression of the mucosal lesions was associated with increased inflammation and decreased differentiation, as indicated by a higher Marsh ranking, higher *IFNG* expression, and lower *TM4SF4* expression (supplementary figure 1). This, in general, inverse relationship between *IFNG* and *TM4SF4* expression in patients is illustrated in figure 1. However, there were also some inconsistencies in individual patients or between biopsy samples. If we consider the design of the study we can anticipate four types of variability, acting at different resolution scales, that could perturb the relationship between the histological classification and the gene expression data: 1) inter-individual variability between similarly diagnosed and Marsh-matched patients; 2) long-range variation between duodenal folds; 3) short-range variations within biopsy-pair distance; and 4) experimental noise in gene expression measurements. We address these issues below.

Table 2. Marsh classifications and *IFNG/TM4SF4* relative gene expression determined from six replicate biopsy samples in each individual

Ind.		Biopsy						Average[1]	SD[2]	CV		
		1	2	3	4	5	6			Biopsy[3]	Exp[4]	Patchy[5]
A	Marsh	0	0	0	0	0	0	1.00	0.00	0.0		0.0
	IFNG	nd[6]	0.57	0.53	0.32	0.47	0.35	0.45	0.11	24.6	17.3	7.3
	TM4SF4	nd	1.32	1.16	1.53	1.02	0.95	1.20	0.24	19.8	5.9	13.9
B	Marsh	0	0	0	0	0	0	1.00	0.00	0.0		0.0
	IFNG	0.64	1.04	0.89	0.72	0.62	1.56	0.91	0.36	39.0	21.6	17.4
	TM4SF4	1.79	1.29	1.72	2.12	2.06	0.97	1.66	0.45	27.1	10.3	16.8
C	Marsh	0	0	0	0	0	0	1.00	0.00	0.0		0.0
	IFNG	0.60	1.11	0.51	0.91	1.13	1.59	0.97	0.40	40.9	19.8	21.1
	TM4SF4	2.22	1.65	1.57	1.57	1.23	2.12	1.73	0.37	21.5	10.6	10.9
D	Marsh	0	0	0	0	0	0	1.00	0.00	0.0		0.0
	IFNG	0.66	0.57	0.84	1.40	0.60	3.67	1.29	1.21	93.6	15.5	78.1
	TM4SF4	1.66	1.65	1.46	2.25	1.79	1.97	1.80	0.28	15.5	11.1	4.4
E	Marsh	0	0	0	0	0	0	1.00	0.00	0.0		0.0
	IFNG	1.27	1.05	2.05	1.54	1.04	1.45	1.40	0.38	27.1	15.9	11.2
	TM4SF4	0.43	0.62	0.36	0.12	0.42	0.49	0.40	0.17	41.1	7.9	33.2
F	Marsh	0	0	0	0	0	0	1.00	0.00	0.0		0.0
	IFNG	1.41	1.61	1.63	2.44	2.68	1.90	1.95	0.51	26.1	16.8	9.3
	TM4SF4	0.59	0.53	0.66	0.31	0.60	0.40	0.51	0.13	26.0	9.1	16.9
G	Marsh	I	I	I	I	I	I	2.00	0.00	0.0		0.0
	IFNG	1.03	0.80	0.65	0.61	0.29	0.44	0.64	0.26	40.9	9.7	31.2
	TM4SF4	0.54	0.50	0.50	0.88	0.64	0.46	0.59	0.16	26.5	10.7	15.8
H	Marsh	I	I	I	I	I	I	2.00	0.00	0.0		0.0
	IFNG	2.09	2.31	2.73	2.40	3.58	3.72	2.81	0.69	24.5	16.7	7.8
	TM4SF4	0.91	0.78	0.61	0.74	0.98	1.41	0.91	0.28	30.8	10.9	19.9
I	Marsh	0	0	0	0	0	0	1.00	0.00	0.0		0.0
	IFNG	0.75	0.41	0.49	0.47	0.33	0.78	0.54	0.19	34.5	13.7	20.8
	TM4SF4	1.00	1.02	1.09	1.06	1.93	0.81	1.15	0.39	34.3	6.5	27.8
J	Marsh	0	0	0	0	0	0	1.00	0.00	0.0		0.0
	IFNG	1.78	1.15	3.58	1.59	1.12	1.27	1.75	0.93	53.4	17.3	36.1
	TM4SF4	1.25	1.59	1.92	1.33	1.39	1.64	1.52	0.25	16.3	6.4	9.9
K	Marsh	I	I	I	I	I	I	2.00	0.00	0.0		0.0
	IFNG	3.29	1.54	2.30	1.19	1.14	1.45	1.82	0.83	45.8	18.0	27.8
	TM4SF4	3.54	1.24	2.20	1.48	1.29	0.92	1.78	0.96	54.3	18.9	35.4
L	Marsh	I	I	I	II	II	II	2.50	0.50	20.0		20.0
	IFNG	10.75	13.20	9.25	6.73	6.55	8.26	9.12	2.54	27.9	9.5	18.3
	TM4SF4	0.70	0.76	0.50	0.64	0.68	0.50	0.63	0.11	17.1	10.0	7.1
M	Marsh	0	0	0	0	0	0	1.00	0.00	0.0		0.0
	IFNG	0.60	0.66	0.60	0.41	0.43	0.53	0.54	0.10	18.7	13.9	4.8
	TM4SF4	0.76	0.74	0.90	0.94	0.82	0.52	0.78	0.15	18.9	10.3	8.6
N	Marsh	I	I	I	I	II	I	2.17	0.37	17.2		17.2
	IFNG	4.99	3.89	2.83	2.27	3.55	3.04	3.43	0.95	27.8	13.3	14.5
	TM4SF4	0.17	0.23	0.09	0.20	0.17	0.18	0.17	0.05	27.7	6.9	20.8
O	Marsh	I	I	I	I	I	I	2.00	0.00	0.0		0.0
	IFNG	3.89	1.77	6.00	5.73	4.94	13.74	6.01	4.08	67.9	14.5	53.4
	TM4SF4	0.56	0.66	0.55	0.53	0.51	0.33	0.52	0.11	20.2	6.8	13.4

Ind.		Biopsy						Average[1]	SD[2]	CV		
		1	2	3	4	5	6			Biopsy[3]	Exp[4]	Patchy[5]
P	Marsh	I	I	I	I	I	I	2.00	0.00	0.0		0.0
	IFNG	13.50	10.64	9.46	9.69	10.13	3.69	9.52	3.21	33.7	12.5	21.2
	TM4SF4	1.28	1.02	1.34	0.80	0.61	1.01	1.01	0.28	27.6	8.8	18.8
Q	Marsh	IIIb	IIIb	IIIb	IIIb	IIIb	IIIb	5.00	0.00	0.0		0.0
	IFNG	8.26	15.43	6.19	1.93	2.48	nd	6.86	5.46	79.6	14.6	65.0
	TM4SF4	0.07	0.04	0.08	0.10	0.10	nd	0.08	0.03	32.5	11.1	21.4
R	Marsh	IIIc	IIIc	IIIb	IIIb	IIIb	IIIb	5.33	0.47	8.8		8.8
	IFNG	22.20	22.05	24.95	nd	19.23	21.27	21.94	2.06	9.4	8.6	0.8
	TM4SF4	0.05	0.03	0.04	nd	0.06	0.03	0.04	0.01	30.1	7.4	22.7
S	Marsh	IIIa	IIIb	IIIc	IIIb	IIIb	IIIb	5.00	0.58	11.5		11.5
	IFNG	18.49	67.61	26.38	26.89	19.05	29.89	31.38	18.32	58.4	13.5	44.9
	TM4SF4	0.27	0.58	0.20	0.37	0.37	0.14	0.32	0.16	48.7	10.4	38.3
T	Marsh	IIIb	IIIc	IIIc	IIIc	IIIc	IIIc	5.83	0.37	6.4		6.4
	IFNG	43.81	49.78	nd	33.71	nd	nd	42.43	8.12	19.1	9.4	9.7
	TM4SF4	0.02	0.04	nd	0.03	nd	nd	0.03	0.01	20.4	7.0	13.4
U	Marsh	I	I	I	II	IIIa	IIIa	2.83	0.90	31.7		31.7
	IFNG	12.86	14.61	6.34	25.78	23.29	15.67	16.43	7.11	43.3	15.3	28.0
	TM4SF4	1.05	0.60	1.26	0.38	0.43	0.23	0.66	0.41	62.1	16.4	45.7
V	Marsh	IIIc	IIIc	IIIc	IIIc	IIIc	IIIc	6.00	0.00	0.0		0.0
	IFNG	31.84	37.42	32.24	47.69	37.56	38.06	37.47	5.72	15.3	9.4	5.9
	TM4SF4	0.02	0.07	0.07	0.05	0.06	0.10	0.06	0.03	41.4	16.9	24.5
W	Marsh	IIIb	IIIc	IIIb	IIIa	IIIb	IIIa	4.83	0.69	14.2		14.2
	IFNG	50.78	45.55	36.58	31.89	24.43	52.77	40.33	11.22	27.8	12.8	15.0
	TM4SF4	0.05	0.06	0.09	0.10	0.14	0.04	0.08	0.04	46.8	13.4	33.4

1 Average. A weighed average was calculated for the Marsh classification using the values indicated in the Material & Methods. Average relative gene expression was calculated from expression data that were normalized to the averaged data obtained from M0 control patients A-F.
2 SD. Standard deviation of the mean.
3 CV_{biopsy}: Coefficient of variation (SD/average x 100) calculated from the six replicate biopsy samples collected from each patient.
4 CV_{exp}: Coefficient of variation calculated from expression values obtained from experiments performed in triplicate, and subsequently averaged over the six replicate biopsy samples.
5 CV_{patchy}: Coefficient of variation of tissue and molecular heterogeneity, calculated as CV_{biopsy} minus $CV_{exp.}$
6 nd. Not determined due to RNA degradation.

Inter-individual variations

Examples of inter-individual expression variation in otherwise closely matched patients can be observed by comparing figure 1 and tables 1 and 2. The non-CD controls A and E, for example, have both a homogeneous M0 mucosa but displayed an inverted expression profile for *IFNG* and *TM4SF4*. Likewise, the M0 remission patients I and J showed opposite *IFNG* regulation, whereas the remission patients J and K, with discordant Marsh lesions, have a nearly identical *IFNG/TM4SF4* profile. This inter-individual variability is difficult to understand but it may be caused by a combination of any of the four noise factors mentioned above. In addition, also cis- and trans-acting genetic factors may exert their influence on gene expression profiles and thus distort the perception of the pathology-driven gene regulation.

Long-range variations

Long-range tissue patchiness was most apparent in RCD patient U who had lesions from three major Marsh classes (I, II, IIIa). This patient partly responded on a GFD (serology negative). Less pronounced heterogeneity, with lesions restricted to the atrophic MIII (a-c) subcategories, were present in three out of four *de novo* CD patients (R, S, and T), and one RCD case (W). Minor heterogeneity (MI-MII) was detectable in one treated CD case (L), and in an untreated DH patient (N) with ambiguous serology. The intra-individual heterogeneity among the six replicate biopsy samples was expressed by the coefficient of variation, CV_{biopsy} (table 2). Comparison of the CV_{biopsy}-values for histology and gene expression revealed no obvious relation. Moreover, patients and controls with a homogeneous histology often also showed a comparable variation in expression (table 2). This suggested that additional factors, beyond tissue patchiness, have an impact on gene expression variation.

Impact of experimental variations

In order to assess the contribution of experimental fluctuations, we compared the CV_{biopsy} for *IFNG* and *TM4SF4* with the average CV_{exp} of the expression measurements that were carried out in triplicate (table 2). On average, the experimental variation accounted for one-third of the gene expression variation (*IFNG:* CV_{biopsy}= 38.2, CV_{exp}= 14.5; *TM4SF4:* CV_{biopsy}= 30.7, CV_{exp}= 10.3). Next we defined the true expression heterogeneity in each individual, CV_{patchy}, as CV_{biopsy} minus CV_{exp}. For simplicity we assumed the tissue patchiness CV_{patchy} to be equal to CV_{biopsy}. Comparison of each patient's tissue CV_{patchy} with its corresponding expression CV_{patchy} showed no correlation with either *IFNG* or *TM4SF4* (table 2, supplementary figure 2). This suggested that long-range expression heterogeneity was present, regardless of the degree of tissue patchiness.

Short-range variations

When patients with tissue patchiness (L, N, R, S, T, U, and W) were examined for gene expression in each pair of juxtaposed biopsies, we could not observe a concordance with the Marsh pattern either. This suggested the presence of a shortrange mucosal heterogeneity (table 2). We conclude that considerable expression heterogeneity is present in the mucosa, both long- and short-range, which appears largely independent of the presence or absence of tissue patchiness.

Histology and gene expression correlations

This limited histology and expression concordance within a single biopsy-pair might affect the relationship Marsh classification – gene expression. To assess to what extent this relationship might be disturbed, we plotted the expression

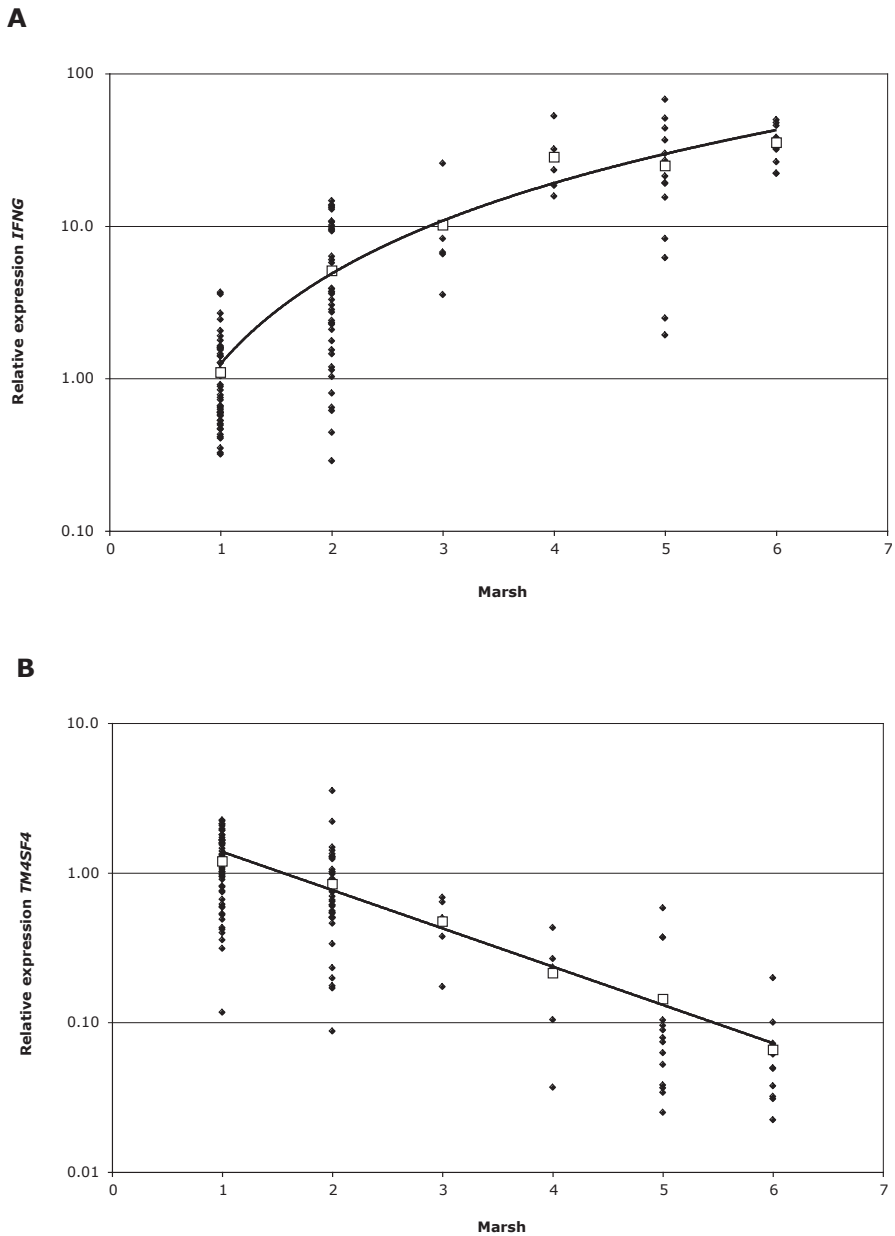


Figure 2. Relative gene expression of *IFNG* (**A**) and *TM4SF4* (**B**) as a function of the Marsh classification in paired biopsy samples, irrespective of patients' diagnosis. Two biopsy samples, harvested as a pair, were separated and used for either gene expression measurement or Marsh typing. Gene expression data (diamonds) are the average of three measurements. The average expression values of the Marsh classes are indicated (squares), and connected by a trend line.

of *IFNG* (figure 2A) and *TM4SF4* (figure 2B) against the Marsh category of the matching biopsy within the pair, irrespective of the patients they originated from. The *IFNG* plot shows an incremental increase from M0 - MIIIa, which then reaches a plateau towards MIIIc. The M0 – MIIIa interval correlates highly with the average *IFNG* expression ($R^2 > 0.98$). The graph for *TM4SF4* shows a consistent decrease in expression over the entire Marsh range, and is also highly correlated with the average gene expression ($R^2 > 0.99$). We conclude that generally, despite the various noise factors present, there is a clearly demonstrable correlation between tissue morphology and gene expression, irrespective of the patients' diagnosis and tissue patchiness.

Since, on average, both *IFNG* and *TM4SF4* expression correlated well with Marsh typing in paired biopsy samples, we also plotted the activity of both genes against each other, regardless of patients' diagnosis or Marsh typing (figure 3A). The *IFNG* and *TM4SF4* expression correlated well ($R^2 = 0.48$), confirming the inverted relationship between inflammation and differentiation that we also observed in individual patients (figure 1).

We argued that, given the opposite responses of *IFNG* and *TM4SF4* to mucosal remodeling, the expression ratio *IFNG/ TM4SF4* would provide a more robust molecular signature to reflect the mucosal impairment. In addition, attaching arbitrary classifiers (1 – 6) to the Marsh categories (M0 – MIIIc) allows a weighted averaged Marsh value for each individual patient that reflects the severity of its lesions and the extent of tissue patchiness. We plotted the averaged *IFNG/TM4SF4* gene expression ratio of each individual patient as a function of its weighted averaged Marsh classifier and observed a strong correlation between these two parameters ($R^2 = 0.89$). This plot could also be applied as a calibration line for inferring gene expression from a weighted Marsh classification, and vice versa. It is noteworthy that individual patients 'move' along this calibration line irrespective of their diagnosis (cf. table 1 and figure 3B), demonstrating that it is mucosal remodeling that is the principal determinant of *IFNG* and *TM4SF4* gene expression.

DISCUSSION

Patchy patterns of tissue lesions have been reported in coeliac disease, both in adults[9], and children[10]. This mosaic distribution of mucosal remodeling is usually assessed by endoscopic forceps grasp sampling of multiple biopsies and histological examination[13]. Alternatively, lesion heterogeneity can be observed using live-imaging techniques like high-magnification immersion chromoscopic duodenoscopy[14]. Generally, the extent of tissue variability in the distal duodenum is limited to neighboring Marsh classes[11], but the underlying mechanism for this phenomenon is not known.

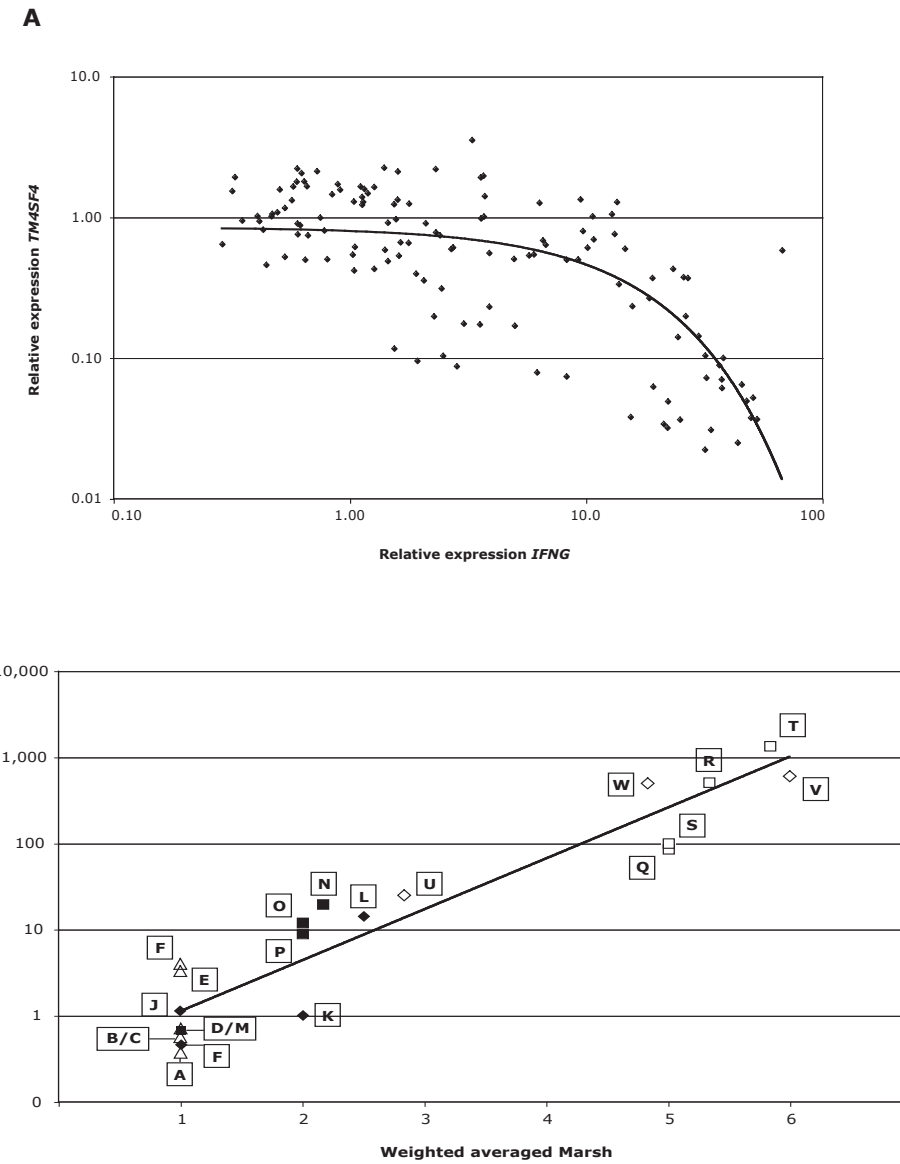


Figure 3. Interdependence of *IFNG, TM4SF4* gene expression, and Marsh stage. **A**) Relationship between *IFNG* and *TM4SF4* gene expression within single biopsy isolates, regardless of Marsh classification and patient's diagnosis. Expression data (diamonds) are the average of three measurements. **B**) Correlation between the weighted averaged Marsh classification of each patient and the ratio of their average *IFNG* and *TM4SF4* expression. Data points are tagged with capital letters, referring to the patients in table 1. Weighting factors from 1–6 were attached to the six Marsh classes M0 – MIIIc, in order. Average gene expression was calculated from six (or less in the case of drop-outs) replicate biopsies per patient. Gene expression data per biopsy are the average of three measurements.

The frequency and extent of tissue patchiness that we observed in our patient group is in line with what has been reported before[11], with exception of one RCD patient. This patient (U) displayed three major Marsh stages, responded partly to a GFD, and may thus represent a more complex and atypical case of tissue mosaicism. In this study we also examined for heterogeneity on the molecular level, using gene expression, and compared it with tissue patchiness. The activity of the genes *IFNG* and *TM4SF4* were used as markers for mucosal inflammation and enterocyte differentiation, respectively. We compared different individuals with similar diagnosis or Marsh lesions, biopsies sampled from different duodenal folds, and biopsies taken as close-spaced pairs from within a single fold. This provided insights into the extent of expression and tissue heterogeneity on different levels of resolution. We also assessed that, on average, one-third of the expression variability can be accounted for by the experimental error in the triplicate measurements. Interestingly, we did not observe a correlation between the gene expression heterogeneity and the tissue patchiness. The coefficients of variation (corrected for the experimental variation) of tissue patchiness and of gene expression were independent. Moreover, the range of expression variation in controls was, in general, similar to that observed in patients, which suggests that expresison heterogeneity is a feature largely independent of tissue patchiness. This was further supported by the lack of correlation between gene expression and Marsh type in paired biopsies from the six replicates taken from each individual. It could also be argued that expression heterogeneity reflects variations in the quality of RNA preparations between the different replicates. However, this would affect the endogenous control gene *GUSB* and the target genes *IFNG* and *TM4SF4* equally. In addition, the PCR fragments generated from these genes are 100 basepairs in size and therefore less sensitive to variations in RNA quality. This type of variation would also generate a constant error in gene expression that would be independent of the actual transcription level. This would result in a diminishing coefficient of variance as function of the expression level, which we did not observe. Together, these observations indicated the presence of a short-range mosaic pattern of gene expression that is largely independent of tissue patchiness. Live imaging of duodenal atrophy showed alternating patches with an area of 5 – 10 villi in size[14]. This level of patchiness is compatible with the scale at which we observed expression heterogeneity.

We propose a model in which a mosaic pattern of gene expression in the duodenum, both in healthy and CD individuals, translates into overt tissue patchiness when a certain inflammatory threshold has been crossed. With further progression of the inflammation the tissue will than become homogeneous, until the next transitional threshold is met and the next temporal phase of patchiness occurs. The expression heterogeneity that we observed may reflect variations in cell

compositions of the tissue patches, or it could point to a differential response of groups of cells to stimuli. It is noteworthy in this respect that, on the cellular level, gene expression variation was recently observed in the aging mouse heart[15].

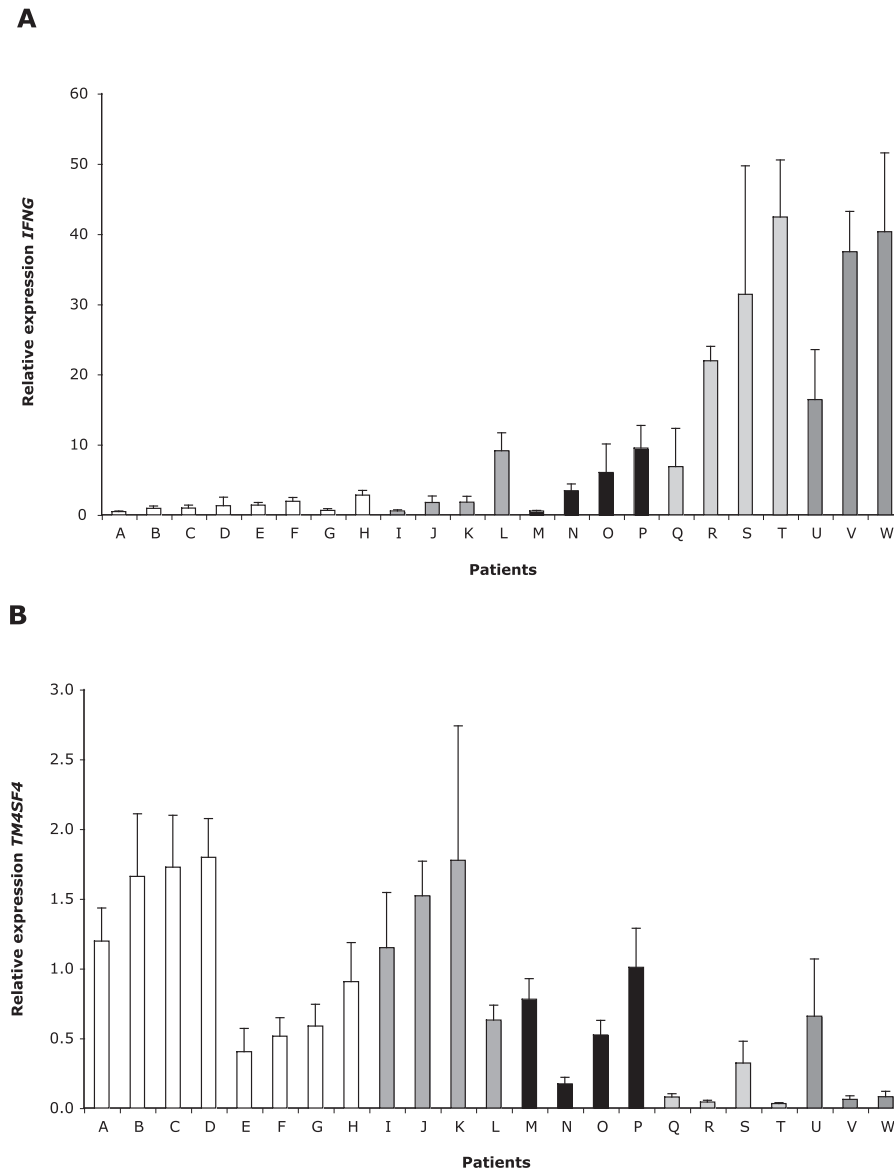
Despite the molecular heterogeneity and other distorting variables there was a clearly demonstrable correlation between the levels of gene expression and disease progression, both on the level of biopsies and individuals. Particularly powerful and applicable was the strong correlation between the averaged *IFNG/TM4SF4* expression ratio and the patients' weighted averaged Marsh typing. That gene expression is predominantly dependent on Marsh typing, rather than disease diagnosis, was convincingly demonstrated by this relationship. In addition, the calibration plot of this inter-dependence can be used to put a Marsh tag on RNA isolates of biopsies that are not, or insufficiently, typed.

Our findings on tissue patchiness are in line with previous studies and will not affect current procedures of histological examination that usually includes at least four biopsy samples[16]. Our *IFNG* and *TM4SF4* expression plots as a function of Marsh-typed biopsies could be considered as prototypes for many other pathology-regulated genes. Many of these genes will show rather subtle changes in expression and will be difficult to identify or pass statistical significance, by using a simple case-control design. However, including the entire Marsh range in such studies will allow the identification of genes with an incremental and consistent increase or decrease in gene expression, like *IFNG* and *TM4SF4*, using standard correlation or linear regression. To best cope with the diverse sources of variation that affect the expression as discussed in this paper, we suggest using 2 – 4 (pooled) biopsies per individual and 10 – 20 individuals for each Marsh classification. Recently, we have implemented this approach successfully to expose molecular pathways of immunity (Diosdado, submitted) and differentiation (Wapenaar, submitted) in the pathology of coeliac disease. In conclusion, the gene expression – tissue pathology relationship we have described here, and its utilization in identifying differentially expressed genes, may be generally applicable to any disease with clearly identifiable stages of tissue remodeling.

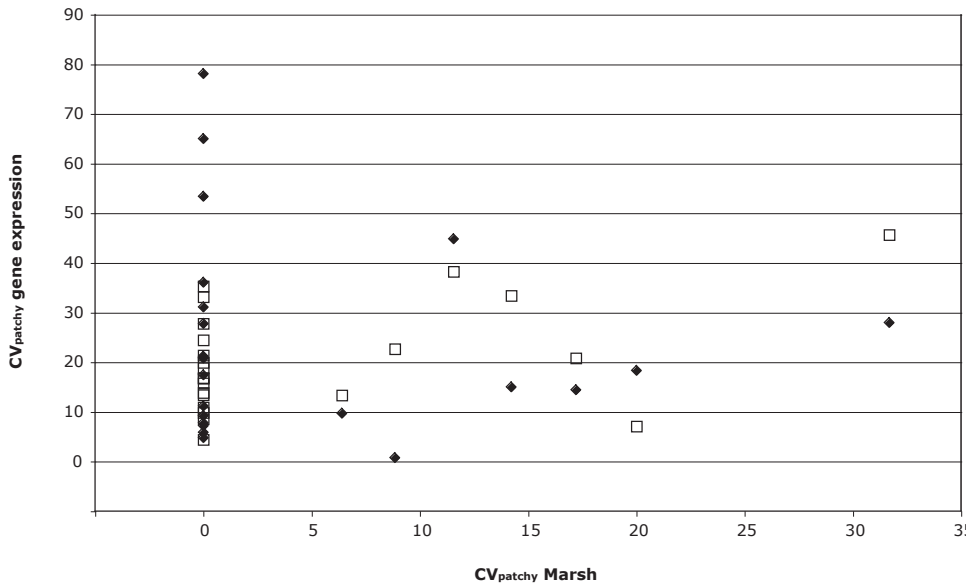
REFERENCES

1. Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. Nat Rev Immunol 2002;2:647-55.
2. van Heel DA, West J. Recent advances in coeliac disease. Gut 2006;55:1037-46.
3. Report of a working group of the United European Gastroenterology Week in Amsterdam. When is a coeliac a coeliac? Eur J Gastroenterol Hepatol 2001;13:1123-28.
4. Rostami K, Kerckhaert J, Tiemessen R, et al. Sensitivity of antiendomysium and antigliadin antibodies in untreated coeliac disease: disappointing in clinical practice. The American journal of gastroenterology 1999;94:888-94.

5. Rostami K, Mulder CJ, Stapel S, et al. Autoantibodies and histogenesis of coeliac disease. Rom J Gastroenterol 2003;12:101-6.
6. Mulder CJ, Bartelsman JF. Case-finding in coeliac disease should be intensified. Best Pract Res Clin Gastroenterol 2005;19:479-86.
7. Wapenaar MC, van Belzen MJ, Fransen JH, et al. The interferon gamma gene in coeliac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility. J Autoimmun 2004;23:183-90.
8. Diosdado B, Wapenaar MC, Franke L, et al. A microarray screen for novel candidate genes in coeliac disease pathogenesis. Gut 2004;53:944-51.
9. Scott BB, Losowsky MS. Patchiness and duodenal-jejunal variation of the mucosal abnormality in coeliac disease and dermatitis herpetiformis. Gut 1976;17:984-92.
10. Bonamico M, Mariani P, Thanasi E, et al. Patchy villous atrophy of the duodenum in childhood coeliac disease. J Pediatr Gastroenterol Nutr 2004;38:204-7.
11. Ravelli A, Bolognini S, Gambarotti M, et al. Variability of histologic lesions in relation to biopsy site in gluten-sensitive enteropathy. The American journal of gastroenterology 2005;100:177-85.
12. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001;25:402-8.
13. Meijer JW, Wahab PJ, Mulder CJ. Small intestinal biopsies in coeliac disease: duodenal or jejunal? Virchows Arch 2003;442:124-8.
14. Hurlstone DP, Sanders DS. High-magnification immersion chromoscopic duodenoscopy permits visualization of patchy atrophy in coeliac disease: an opportunity to target biopsies of abnormal mucosa. Gastrointest Endosc 2003;58:815-6.
15. Bahar R, Hartmann CH, Rodriguez KA, et al. Increased cell-to-cell variation in gene expression in ageing mouse heart. Nature 2006;441:1011-4.
16. Branski D, Faber J, Freier S, et al. Histologic evaluation of endoscopic versus suction biopsies of small intestinal mucosae in children with and without coeliac disease. J Pediatr Gastroenterol Nutr 1998;27:6-11.



Supplementary figure 1. Relative gene expression of *IFNG* (**A**) and *TM4SF4* (**B**) in 23 patients and controls. Individuals are indicated with capital letters A-W (see table 1 for details). Vertical bars represent the relative expression calculated from the average of six replicate biopsy samples (see table 2). Error bars indicate the standard deviation. Each biopsy sample was tested in triplicate. Expression values were normalized to the averaged expression of the controls A-F.



Supplementary figure 2. The relation between the CV_{patchy}-values of tissue patchiness and expression heterogeneity for *IFNG* (filled diamonds) and *TM4SF4* (open squares) in 23 individuals.

Chapter 4

A microarray screen for novel candidate genes in coeliac disease pathogenesis

B Diosdado, MC Wapenaar, L Franke, KJ Duran,
MJ Goerres, M Hadithi, JBA Crusius, JWR Meijer,
DJ Duggan, CJJ Mulder, FCP Holstege, and C Wijmenga

*Gut 2004; **53**: 944-51*

(reproduced with permission of the BMJ Publishing Group)

SMALL INTESTINE

A microarray screen for novel candidate genes in coeliac disease pathogenesis

B Diosdado, M C Wapenaar, L Franke, K J Duran, M J Goerres, M Hadithi, J B A Crusius, J W R Meijer, D J Duggan, C J J Mulder, F C P Holstege, C Wijmenga

Gut 2004;**53**:944–951. doi: 10.1136/gut.2003.018374

See end of article for authors' affiliations

Correspondence to: Professor C Wijmenga, Department of Biomedical Genetics, Stratenum 2. 117, University Medical Centre Utrecht, PO Box 80030, 3508 TA Utrecht, the Netherlands; t.n.wijmenga@med.uu.nl

Background and aims: The causative molecular pathways underlying the pathogenesis of coeliac disease are poorly understood. To unravel novel aspects of disease pathogenesis, we used microarrays to determine changes in gene expression of duodenal biopsies.

Methods: cDNA microarrays representing 19 200 genes were used to compare gene expression profiles of duodenal biopsies from 15 coeliac disease patients with villous atrophy (Marsh III) and seven control individuals with normal biopsies (Marsh 0). In addition, the specific effect of gluten was studied by comparing the expression profiles of Marsh III lesions of seven patients exposed to gluten with four patients on a gluten free diet.

Results: Comparing Marsh III with Marsh 0 lesions identified 109 genes that differed significantly ($p<0.001$) in expression levels between patients and controls. A large number of these genes have functions in proliferation and differentiation pathways and might be important for correct development of crypt-villous units. Alterations in these pathways may lead to the characteristic hyperplasia and villous atrophy seen in coeliac disease. The analyses also revealed 120 differentially expressed genes ($p<0.005$) when comparing patients on a gluten free diet with those exposed to gluten. These genes further strengthen our observation of increased cell proliferation in the presence of gluten.

Conclusions: Our study provides new candidate genes in the pathogenesis of coeliac disease. Based on our results, we hypothesise that villous atrophy in coeliac disease patients is due to failure in cell differentiation. These genes are involved in pathways not previously implicated in coeliac disease pathogenesis and they may provide new targets for therapy.

Coeliac disease (OMIM 212750) is a chronic inflammatory enteropathy caused by lifelong intolerance to gluten in genetically predisposed individuals. When gluten, the main protein present in wheat, rye, and barley,[1] is ingested by individuals expressing human leucocyte antigen (HLA)-DQ2 and/or -DQ8 heterodimers,[2] it provokes a misdirected immune response in the small intestine where gluten is absorbed. This immune response leads to a series of histological changes resulting in lymphocytosis, crypt hyperplasia, and villous atrophy. These histological abnormalities, which were documented by Marsh in 1992,[3] are characteristic of this disease and explain some of the clinical symptoms of coeliac disease patients. Treatment of coeliac disease consists of a lifelong gluten free diet, resulting in clinical recovery and histological normalisation of the intestine.

Coeliac disease is a multifactorial disorder for which both genetic and environmental factors are required for the disease phenotype to develop.[4] The molecular pathways responsible for the disease pathogenesis are currently only partly defined.[5] So far only one genetic factor has been identified, the HLA-DQA and -DQB gene cluster on chromosome 6.[2] Although many attempts are underway to identify the other genetic factors involved in coeliac disease, traditional genetic mapping studies face serious limitations in identifying the full repertoire of susceptibility genes due to the small contribution of each individual susceptibility gene and the complex interplay between genetic and environmental factors. Genomics technologies, such as cDNA microarrays, are expected to provide additional insight into the molecular and cellular aspects of diseases.

Coeliac disease is a unique disease as the environmental factor, gluten, that triggers the disease is known and can be easily manipulated in vivo as well as in vitro. Moreover, the site of the lesion is easily accessible and small intestinal biopsies need to be taken from coeliac disease patients as part of the standard diagnostic procedure. In order to expand our understanding of the pathogenesis of coeliac disease, we set out to perform cDNA microarray analysis on a set of well characterised duodenum biopsies from coeliac disease patients with classic histopathology (Marsh III (MIII)) and control individuals (Marsh 0 (M0)). Microarray analysis is a powerful technique that allows the study of the level and pattern of expression of thousands of genes simultaneously.[6]

We first thought that comparing MIII with M0 biopsies might lead to the discovery of genes involved in the immune response to gluten and the long term tissue destruction seen in the small intestine of coeliac disease patients. Subsequently, we investigated the molecular changes that occur in MIII biopsies by comparing a collection of MIII biopsies from individuals who were on a gluten free diet with those on a gluten containing diet (that is, exposed to gluten).

We report two sets of genes that have not previously been associated with the pathogenesis of coeliac disease. Our results imply a role for novel candidate genes involved in the maintenance of the intestinal villi. As some of these genes map to chromosomal regions implicated in genetic mapping studies, they may represent possible causal candidate genes.

The combination of well characterised intestinal biopsies and cDNA microarrays is a unique method of studying the molecular and cellular events taking place in the initiation

Abbreviations: HLA, human leucocyte antigen; M, Marsh; Th, T helper; RT-PCR, reverse transcription-polymerase chain reaction; IFN, interferon; IL, interleukin; APC, antigen presenting cells

and further progression towards the mucosal transformation seen in coeliac disease. It has not yet been applied to this disease.

MATERIALS AND METHODS

Patients

Intestinal biopsies from 15 coeliac disease patients were included in the study. All biopsies showed MIII histology and were evaluated by one pathologist (JWRM) according to the modified UEGW criteria.[7]

Seven patients (patients 1–7, table 1) suspected of having coeliac disease were on a gluten containing diet when MIII villous atrophy was histologically confirmed.

Another four coeliac disease patients (patients 8–11, table 1) were on a gluten free diet for a year and reported total clinical recovery although their intestines still showed MIII characteristics. A slight histological improvement was reported. In addition, biopsies were taken from the last group of four patients (patients 12–15, table 1), who were refractory coeliac disease type I patients.[8] Biopsies of these four refractory coeliac disease type I patients showed MIII histology and no clinical improvement[8] despite the fact that they were on a strict gluten free diet. A team of dieticians monitored compliance to the diet in all patients on a gluten free diet.

In addition, duodenum biopsies from seven individuals who had an endoscopic examination for other reasons were used as control samples. The histology of these tissue biopsies was completely normal (M0). Characteristics of the coeliac disease patients and controls included in this study are summarised in table 1.

The Medical Ethical Committee of the University Medical Centre of Utrecht approved the study. All patients, or the parents of paediatric patients, included in the study gave written informed consent.

Biopsy sampling and RNA isolation

For each individual, two to three biopsies (15–20 mg) were taken from the proximal duodenum by spike forceps endoscopy. Fresh tissue samples were snap frozen and stored in liquid nitrogen. Frozen biopsies were homogenised in TRIzol with glass beads of 1 mm diameter using a Mini-BeadBeater (BioSpec Products, Inc, Bartlesville, Oklahoma, USA), and total RNA was isolated using TRIzol (Gibco/Life Technologies, Rockville, Maryland, USA) following the manufacturer's protocol. The quality and quantity of the RNA samples was determined using a 2100 Agilent Bioanalyzer (Agilent Technologies, Palo Alto, California, USA). Biopsies yielded an average of 3.6 µg of RNA/mg of tissue.

Microarray hybridisation

The seven controls and 15 coeliac disease samples were analysed by hybridisation screening of cDNA microarrays obtained from the University Health Network of Toronto, Ontario (Canada). Two different releases of slide sets, 19k2 and 19k3, containing 19 200 genes printed in duplicate on two glass slides were used for the experiments.

First strand cDNA was prepared from 10 µg of total RNA from biopsies and labelled with Cy3, as described by Van de Peppel and colleagues.[9] A surgical specimen from the small bowel was used to generate a Cy5 labelled cDNA probe for use as a reference in all coeliac disease and control hybridisations. Between 200 and 300 ng of labelled cDNA from the biopsies and reference tissues were used for a single set of slides for overnight hybridisation. Slides were scanned in a ScanArray 4000 XL (Packard BioScience, Boston, Massachusetts, USA).

An MIAME compliant[10] downloadable dataset and full details of the protocols are available at http://humgen.med.uu.nl/publications/CD/Microarrays/Diosdado_1.html or http://www.ebi.ac.uk/arrayexpress/, experiment accession No E-MEXP-42.

Image and data analysis

The scanned images were subjected to image analysis using the Dearray suite[11] for IPLab SPECTRUM software (Scanalytics, Inc., Fairfax, Virginia, USA). The software, in a

Table 1 Characteristics of coeliac disease, refractory gluten free diet type I patients, and controls included in the study

Sample	Age/Sex	Diagnosis	Biopsy status‡	GFD¶	HLA typing	Slide type§
1	18/F	CD patient*	MIIIb	No	DQ2+/DQ8−	19k2
2	3/M	CD patient	MIIIb	No	DQ2+/DQ8−	19k3
3	54/F	CD patient	MIIIc	No	DQ2+/DQ8−	19k3
4	33/F	CD patient	MIIIa	No	DQ2+/DQ8−	19k3
5	2/F	CD patient	MIIIc	No	DQ2+/DQ8−	19k3
6	79/F	CD patient	MIIIb	No	DQ2+/DQ8−	19k3
7	38/M	CD patient	MIIIa	No	DQ2+/DQ8−	19k2
8	61/M	CD patient	MIIIb	Yes	DQ2+/DQ8−	19k3
9	40/M	CD patient	MIIIb	Yes	DQ2+/DQ8−	19k3
10	61/F	CD patient	MIIIa	Yes	DQ2+/DQ8−	19k3
11	70/M	CD patient	MIIIa	Yes	DQ2+/DQ8−	19k3
12	51/F	RCD type I patient†	MIIIa	Yes	DQ2+/DQ8−	19k3
13	29/F	RCD type I patient	MIIIb-c	Yes	DQ2+/DQ8+	19k3
14	84/F	RCD type I patient	MIIIa	Yes	DQ2+/DQ8−	19k3
15	72/F	RCD type I patient	MIIIa	Yes	DQ2-/DQ8−	19k3
16	32/F	Control	M0	No	DQ2+/DQ8+	19k2
17	40/F	Control	M0	No	DQ2-/DQ8−	19k3
18	51/F	Control	M0	No	n/d	19k3
19	60/F	Control	M0	No	n/d	19k3
20	36/F	Control	M0	No	DQ2-/DQ8−	19k2
21	32/F	Control	M0	No	n/d	19k2
22	33/F	Control	M0	No	DQ2-/DQ8+	19k2

*CD patient, coeliac disease patient; †RCD patient, refractory gluten free diet type I patient.
‡Histology was reviewed by the same pathologist (JWRM). Biopsies were classified according to the UEGW criteria.
¶GFD, gluten free diet.
§Two releases of microarray slides, 19k2 and 19k3, from the University Health Network, Toronto, Ontario (Canada) were used for hybridisations.

semi-automatic manner, identifies the fluorescent spots, subtracts the local background, and determines a quality score for each spot based on the spot's intensity, size, local background, and uniformity of intensity within the spot.[12] Based on these parameters, a quality score is assigned to each individual spot. All spots with a quality score >0.01 were selected for further processing.

To correct for differences in the efficiencies between the Cy3 and Cy5 channel, a method of global normalisation was applied. The quantified signal intensity for the entire array in both the Cy3 and Cy5 channels was averaged and equalised by applying a normalisation factor. Subsequently, this normalisation correction was applied to each individual spot and the red to green ratio was calculated.[12] As genes are spotted in duplicate, the average of the ^{10}log of the signal of each of the two copies of the same gene was calculated only when the quality score for both genes was >0.01. When only one copy of the gene had a quality score >0.01, only the level of expression of that gene copy was used.

Data analysis was performed with the GeneSpring package, version 4.2.1 (Silicon Genetics, Redwood City, California, USA). Genes whose expression was significantly different between two groups of biopsy samples were selected from the genes present on the slides by applying a Welch *t* test. Cluster analysis was performed using a supervised cluster algorithm.

Gene analysis

To further define the biological function of the selected genes, a homemade Java tool database was developed (Franke *et al*, in press) for data storage, gene classification, and gene analysis. This database also contains information on selected genes such as GeneBank accession No, Locus Link ID, chromosomal location, Emsembl ID, Unigene information, Gene Ontology, and GeneCards ID.

Data validation by real time RT-PCR

A selection of genes that showed altered expression in the microarray analysis was re-examined by real time reverse transcription-polymerase chain reaction (RT-PCR) to validate the changes observed in an independent manner. Firstly, cDNA was generated from 1 µg of total RNA using the High Capacity cDNA Archive Kit. PCR cycling was performed on a 7900HT Sequence Detection System in 25 µl SYBR Green PCR Master Mix using 25 ng of reverse transcribed RNA. Target genes were tested for using Assay-on-Demand Gene Expression products. The GUSB gene was used as an endogenous reference to control for expression independent sample to sample variability. Relative expression was determined from the obtained Ct values and the $2^{-\Delta\Delta Ct}$ method.[13] All equipment and reagents were purchased from Applied Biosystems (Foster City, California, USA) and used according to their protocols. Six genes (ALDOB, IL2RB, PDE7B, TM4SF4, TXN, and TYK2), together with the GUSB reference, were tested in duplicate on pooled cDNA from M0 normal controls (n = 16) and MIII coeliac patients (n = 15). The pooling method was used only after we had assessed that the mean of the data obtained with ALDOB tested on individual samples was similar to that of the pooled samples (table 2). This pooling method was in agreement with our experiences with various other genes not directly related to this study.

Table 2 Relative levels of average gene expression levels in Marsh III coeliac patients versus normal controls, as determined by microarray hybridisation and validation by real time reverse transcription-polymerase chain reaction

Gene	Microarray	RT-PCR
ALDOB*	0.51	0.23
ALDOB†	0.51	0.28
IL2RB2	1.36	1.28
PDE7B	0.64	0.94
TM4SF4	0.50	0.12
TXN	1.26	1.33
TYK2	1.30	0.99

Real time RT-PCR data for ALDOB are indicated as the mean of individual samples* and as pooled samples† of normal controls (n = 16) and Marsh III coeliac patients (n = 15). The other genes were tested only on pools.

RESULTS

Identification of genes differentially expressed in MIII biopsies versus M0 biopsies

In order to identify genes involved in the pathogenesis and therefore in the inflammatory and immune response evoked by gluten and leading to the tissue destruction observed in the duodenal biopsies of coeliac disease patients, the expression profiles of 15 MIII biopsies from patients were compared with seven M0 biopsies from control individuals. All biopsy RNA samples from MIII coeliac disease patients, including four refractory coeliac disease type I samples, were hybridised onto cDNA microarrays. A Welch *t* test with a threshold p value of <0.001 revealed no differentially expressed genes when comparing MIII refractory coeliac disease type I biopsies with MIII coeliac disease biopsies (data not shown). As these two groups could not be distinguished based on their expression levels, they were treated as a single group based on their histological characteristics.

From all 19 200 genes present on the arrays, 10 674 had sufficient data for comparison (that is, genes with a quality score >0.01). A Welch *t* test with a threshold of p<0.001 was applied to the data set and 109 genes were identified to be differentially expressed between MIII and M0 biopsy samples. Approximately 11 genes were expected by chance alone. A distance measurement was used to define the similarity of the expression profiles for both the 22 samples and the 109 genes, and was depicted in a two dimensional hierarchical dendrogram (fig 1A). The hierarchical cluster analysis revealed two branches clearly separating the M0 and MIII biopsies, with the exception of control No 19 that clustered with the group of patient samples.

Of these 109 genes, 76 (69.7%) had an increased level of expression in coeliac disease patients versus controls, and 33 genes (30.3%) had a decreased level of expression in coeliac disease patients versus controls. The molecular function was known or could be predicted for 46 of these 109 genes (table 3).

Some of the differentially expressed genes result from the histological changes but a significant number of them result from the inflammatory response seen in MIII lesions of coeliac disease patients. The results from these experiments are consistent with the widely accepted T helper (Th)1 response.[14] Although this observation in itself adds little new insight, it does replicate previous independent findings. Hence these observations support our results and thereby validate microarray technology as a useful tool in providing new insight into the pathogenesis of coeliac disease.

Although the key cytokines interferon (IFN)-γ and tumour necrosis factor α were not present on our slides, real time RT-PCR of the *IFNG* gene showed extremely high levels of expression of ~30-fold in MIII biopsies versus M0 biopsies. Stimulation of the interleukin (IL)-2 signalling pathway was further suggested by increased expression of the IL-2 receptor beta (*IL-2RB*) gene which was validated by real-time RT-PCR (table 2), a member of the RAS oncogene family (*RAB1B*), chromosome 20 open reading frame 64 gene (*C20orf64*), and the RAP1, GTP-GDP dissociation stimulator 1 gene (*RAP1GDS1*). Furthermore, the nuclear factor κB pathway seems to be induced, as suggested by upregulation

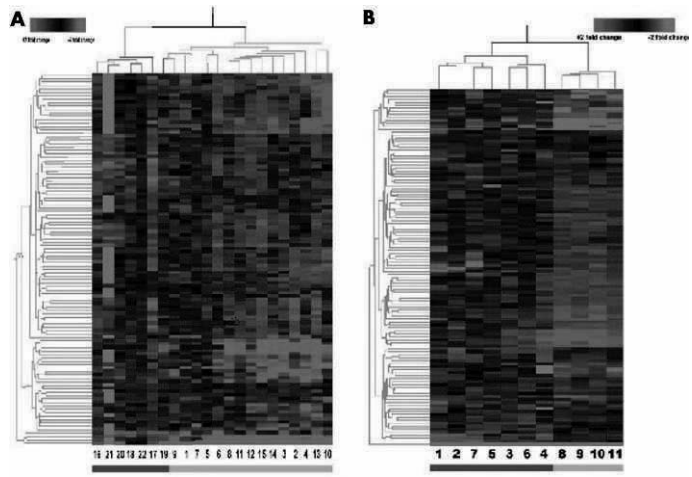


Figure 1 (A) Hierarchical clustering dendrogram of duodenal genes from Marsh III (MIII) and Marsh 0 (M0) biopsies. Clustering of 109 genes across 22 samples clusters the seven control samples (blue bar) separately from the 15 coeliac disease patients (orange bar). Each column represents a coeliac disease (MIII) or a control (M0) sample and each row represents an individual gene. For each gene, a green signal represents underexpression, black signals denote similarly expressed genes, a red signal represents overexpressed genes, and grey signals denote missing data. Information on these genes can be found in table 3 and on the website http://humgen.med.uu.nl/publications/CD/Microarrays/Diosdado_1.html. (B) Hierarchical clustering dendrogram of duodenal genes from MIII biopsies with or without exposures to gluten. Clustering of 120 genes across 11 samples clusters the seven samples of coeliac disease patients following a gluten containing diet (blue bar) separately from the four coeliac disease patients on a gluten free diet (orange bar). Each column represents a sample and each row represents an individual gene. For each gene, a green signal represents underexpression, black signals denote similarly expressed genes, a red signal represents overexpressed genes, and grey signals denote missing data. Information on these genes can be found in table 3 and on the website http://humgen.med.uu.nl/publications/CD/Microarrays/Diosdado_1.html.

of the thioredoxin (*TXN*) gene and the lymphocyte cytosolic protein 1 (*LCP1*) gene. Genes involved in complementary functions, such as complement component 8 β polypeptide (*C8B*) and the H factor 1 (*HF1*), were downregulated, suggesting depletion of complement components.

Upregulation of both the *TXN* and the macrophage scavenger receptor 1 (*MSR1*) genes in MIII versus M0 provides evidence for the presence of active macrophages at the lesion site. MSR1 is expressed by antigen presenting cells (APC) and mediates activation of T cells and promotes adhesion of activated B cells.[15 16] It has also been suggested that MSR class A is involved in the breakdown of the T cell self tolerance in mice.[17 18] TRX enhances the immune response by facilitating both a microenvironment for APCs and for T cell interaction. It has been further proposed that TRX acts as coadjuvant in cytokine mediated lymphocyte proliferation between the APC which presents the gluten to the CD4+ T cells.[19] Upregulation of *TRX* was validated by real time RT-PCR (table 2).

Many genes coding for proteins involved in lipid metabolism and cholesterol homeostasis, such as ATP binding cassette, subfamily A, member 7 (*ABCA7*), apolipoprotein A-II (*APOA2*), and diaphorase (NADH) (*DIA1*), were upregulated in MIII biopsies whereas hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid 1 (*HSD3B1*) was downregulated. Expression studies have shown that γδ T cells differentially express a large number of genes involved in lipid and cholesterol homeostasis, implying that such cells may be involved in these mechanisms.[20] One of the histological characteristics of MIII coeliac disease biopsies is the presence of high numbers of intraepithelial γδ T cells.[3] Hence these results may reflect the increased number of γδ T cells in the intestines of coeliac disease patients. Alternatively, increased expression of lipid and cholesterol genes may point to increased biosynthesis of lipid bilayers. In both cases, upregulation of these genes may be due to an increase in the number of cells and not to upregulation ''per cell'' of these genes.

Downregulation of the transmembrane 4 superfamily member 4 gene (TM4SF4), which by real time RT-PCR turned out to be down-regulated eightfold (table 2), and upregulation of the retinol binding protein 4 gene (*RBP4*), are both consistent with increased proliferation of cells in MIII biopsies. Also, upregulation of annexin A6 (*ANXA6*), synaptopodin (*KIAA1029*), and the solute carrier family 9 isoform 3 regulatory factor 1 (*SLC9A3R1*) genes suggests an increased number of cells in early differentiation.

In summary, the differentially expressed genes point towards an increased number of γδ T cells and macrophages at the lesion site, a Th1 response, and increased cell proliferation.

Differential gene expression in MIII biopsies in response to gluten withdrawal

To identify genes specifically involved in the response to the environmental trigger gluten, MIII biopsies were subsequently divided into two groups. Four MIII biopsies from coeliac disease patients who were on a gluten free diet for one year were compared with seven MIII biopsies from patients who were following a gluten containing diet at the time of biopsy. The four biopsies of the refractory coeliac disease type I patients were excluded from this analysis as these patients do not show clinical and histological improvement on gluten withdrawal, and so different molecular events may be occurring in their biopsies.

Table 3 Annotated genes differentially expressed in Marsh III (MIII) versus Marsh 0 (M0), grouped according to their function*

Gene symbol	Gene name	LocusLink ID	Chromosome location	Ratio MIII/M0†
Immune response related				
C8B	Complement component 8, beta polypeptide	732	1p32.1	0.67
HF1	H factor 1 (complement)	3075	1q31.3	0.80
MSR1	Macrophage scavenger receptor 1	4481	8p22	1.25
TXN	Thioredoxin	7295	9q31.3	1.26
IL2RB	Interleukin 2 receptor β	3560	22q13.1	1.36
Chemokines, cytokines, and growth factors				
BTN2A1‡	Butyrophilin, BTF1 precursor	11120	6p22.2	1.18
Inflammatory mediators				
PRG1	Proteoglycan 1, secretory granule	5552	10q22.1	1.21
Cancer related genes				
DKFZP586A011	Cervical cancer 1 proto oncogene	25875	12q13.13	0.72
C20orf64	Chromosome 20 open reading frame 64	112858	20q13.12	1.30
Structural related genes				
TM4SF4	Transmembrane 4 superfamily member 4	7104	3q25.1	0.50
IGFBP7	Insulin-like growth factor binding protein 7	3490	4q12	0.73
EPB42	Erythrocyte membrane protein band 4.2	2038	15q15.2	0.79
BTN2A1‡	Butyrophilin, BTF1 precursor	11120	6p22.2	1.18
SLC9A3R1	Solute carrier family 9, isoform 3 regulatory factor 1	9368	17q25.1	1.20
RBP4	Retinol binding protein 4, interstitial	5950	10q23.33	1.24
KIAA1029	Synaptopodin	11346	5q33.1	1.43
Signal transduction				
PPP3CB	Calcineurin A beta	5532	10q22.2	0.76
RAP1GDS1	RAP1, GTP-GDP dissociation stimulator 1	5910	4q23	1.22
RAB1B	RAB1B, member RAS oncogene family	81876	11q13.3	1.23
ANXA6	Annexin A6	309	5q33.1	1.24
LCP1	Lymphocyte cytosolic protein 1	3936	13q14.13	1.30
TYK2	Tyrosine kinase 2	7297	19p13.2	1.30
JFC1	NADPH oxidase related, C2 domain containing protein	84958	1p35.2	1.34
Metabolic pathways and ion transport mediators				
ALDOB	Aldolase B, fructose bisphosphate	229	9q31.1	0.51
PDE7B	Phosphodiesterase 7B	27115	6q23.3	0.64
FLJ12899	Pantothenate kinase 3	79646	5q34	0.71
HSD3B1	Hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid-1	3283	1p12	0.82
ABCA7	ATP binding cassette, subfamily A, member 7	10347	19p13.3	1.18
APOA2	Apolipoprotein A-II	336	1q23.3	1.21
SRI	Sorcin	6717	7q21.12	1.23
GYG2	Glycogenin 2	8908	Xp22.33	1.27
SMUG1	Single strand selective monofunctional uracil DNA glycosylase	23583	12q13.13	1.29
MRPS22	Mitochondrial ribosomal protein S22	56945	3q23	1.29
KCNQ3	Potassium voltage gated channel, KQT-like subfamily, member 3	3786	8q24.22	1.30
ATP1A2	ATPase, Na+/K+ transporting, alpha 2 (+) polypeptide	477	1q23.2	1.30
DIA1	Diaphorase (NADH)	1727	22q13.2	1.30
Cell cycle regulators and transcription factors				
HBP1	HMG-box containing protein 1	26959	7q22.3	0.83
KHDRBS1	Sam 68	10657	1p35.1	1.11
NAP1L4	Nucleosome assembly protein 1-like 4	4676	11p15.4	1.27
C20orf64	Chromosome 20 open reading frame 64	112858	20q13.12	1.30
C21orf66	Chromosome 21 open reading frame 66	94104	21q22.11	1.38
Cell-cell signalling				
ECM2	Extracellular matrix protein 2	1842	9q22.31	1.24
RIMS1	Regulating synaptic membrane exocytosis 1	22999	6q13	1.29
GRM3	Glutamate receptor, metabotropic 3	2913	7q21.12	1.37
Protein synthesis				
EIF4EBP3	Eukaryotic translation initiation factor 4E binding protein 3	8637	5q31.3	0.72
EIF2B3	Eukaryotic translation initiation factor 2B, subunit 3	8891	1p34.1	1.27

*All of the genes can be found at http://humgen.med.uu.nl/publications/CD/Microarrays/Diosdado_1.html.
†Ratio of the expression level for each individual gene when comparing MIII coeliac disease biopsies with M0 biopsies. Genes with values <1 are downregulated; genes with values >1 are upregulated.
‡Genes located in more than one functional group.

A Welch *t* test was applied to 11 938 genes that had sufficient data for comparison; 120 genes showed differential expression at a threshold p value <0.005 (table 4). We are aware of the limitations of this small number of samples. This experimental design is therefore less robust than the MIII versus M0 comparison as the number of genes to be expected by change alone would be 60. Although the data should be interpreted with care, the general picture that emerges from the data may give inroads into the effect of gluten on the intestine. A two dimensional hierarchical cluster showed a dendrogram tree (fig 1B) in which the four patients on a gluten free diet clearly clustered separately from the seven patients on a gluten containing diet.

Table 4 Annotated genes differentially expressed in Marsh III (MIII) biopsies with or without gluten, grouped according to functional pathways*

Gene symbol	Gene name	LocusLink ID	Chromosome location	Ratio gluten−/gluten+†
HLA and immune function genes				
C5R1	Complement component 5 receptor 1	728	19q13.32	1.14
CD79B	CD79B antigen	974	17q23.3	1.14
MAP4K4	Mitogen activated protein kinase kinase kinase kinase 4	9448	2q11.2	1.31
ITGA4	Integrin, alpha 4	3676	2q31.3	1.39
Chemokines, cytokines, and growth factors				
None				
Inflammatory mediators				
PDE4D	Phosphodiesterase 4D, cAMP specific	5144	5q11.2	0.67
AUH‡	AU specific RNA binding protein	549	9q22.31	0.86
Cancer related genes				
DEPC-1	Prostate cancer antigen-1	221120	11p12	0.79
M17S2	Membrane component, chromosome 17, surface marker 2	4077	17q21.31	1.24
RABL2B	RAB-like protein 2B	11158	22q13.33	1.35
Structure related genes				
ADAMTS9	A disintegrin-like and metalloprotease with thrombospondin type 1 motif, 9	56999	3p14.1	0.74
ADRM1	Adhesion regulating molecule 1	11047	20q13.33	0.82
KL	Klotho	9365	13q13.1	1.11
FN1	Fibronectin 1	2335	2q35	1.30
Metabolic pathways and ion transport mediators				
PREP	Prolyl endopeptidase	5550	6q21	0.56
PLCB4	Phospholipase C, beta 4	5332	20p12.2	0.79
AUH‡	AU specific RNA binding protein	549	9q22.31	0.86
RYR3	Ryanodine receptor 3	6263	15q14	1.22
IDH3A	Isocitrate dehydrogenase 3, alpha subunit	3419	15q25.1	1.21
BG1	Lipidosin	23205	15q23	1.21
SLC25A4	Adenine nucleotide translocator 1	291	4q35.1	1.25
CPT1A	Carnitine palmitoyltransferase I, liver	1374	11q13.3	1.31
Cell cycle regulators and transcription factors				
LOC51174	Delta-tubulin	51174	17q23.2	0.54
KIF5B	Kinesin family member 5B	3799	10p11.22	0.76
CCNG1	cyclin G1	900	5q34	0.77
NT5C2	5′-nucleotidase, cytosolic II	22978	10q24.32	0.83
ZNF26	Zinc finger protein 26	7574	12q24.33	1.19
POLR2C	Polymerase II, RNA subunit C	5432	16q13	1.19
GTF2H1	General transcription factor IIH, polypeptide 1	2965	11p15.1	1.24
MSI2	Musashi homologue 2 (Drosophila)	124540	17q23.2	1.28
ZNF317	Zinc finger protein 317	57693	19p13.2	1.32
CDC25A	Cell division cycle 25A	993	3p21.31	1.46
PPP2R3A	Protein phosphatase 2 (formerly 2A), regulatory subunit B′, alpha	5523	3q22.3	1.49
POLD3	DNA polymerase delta subunit 3	10714	11q13.4	1.58
Signal transduction				
SYPL	Synaptophysin-like protein	6856	7q22.3	0.79
WAC	WW domain containing adapter with a coiled-coil region	51322	10p12.1	0.83
HNT	Neurotrimin	50863		0.86
RGS16	Regulator of G protein signalling 16	6004	1q25.3	1.23
PSCD3	Pleckstrin homology, Sec7, and coiled/coil domains 3	9265	7p22.1	1.33
CSPG6	Chondroitin sulphate proteoglycan 6 (bamacan)	9126	10q25.2	1.36
Others				
	Pro-neuregulin-3 precursor		10q23.1	0.69
HSPA6	Heat shock 70 kDa protein 6	3310	1q23.3	0.76
SARM	Sterile alpha and HEAT/Armadillo motif protein, orthologue of Drosophila	23098	17q11.2	1.26
AMSH	Associated molecule with the SH3 domain of STAM	10617	2p13.1	1.29
PERQ1	Postmeiotic segregation increased 2-like 12	64599	7q22.1	1.59
HTF9C	Hpall tiny fragments locus 9C	27037	22q11.21	1.61

*All of the genes can be found at http://humgen.med.uu.nl/publications/CD/Microarrays/Diosdado_1.html.
†Ratio of the expression level for each individual gene between biopsies from MIII coeliac disease patients on a gluten free diet (gluten−) and MIII coeliac disease patients following a gluten containing diet (gluten+). Genes with values <1 are downregulated; genes with values >1 are upregulated.
‡Genes located in more than one functional group.

Of these 120 genes, 46 (38.3%) were downregulated and 75 (62.5%) were upregulated in patients on gluten versus coeliac disease patients refraining from gluten. Functional annotation was available for 42 of the 120 genes. Interestingly, the majority of the differentially expressed genes suggest transcription is enhanced and mitotic activity is suppressed in MIII coeliac disease patients who refrain from gluten.

Several of the genes were related to cell division and cell cycle processes. The cyclin G1 (*CCNG1*) gene was downregulated, indicating that less cells go from the G1 to S phase. Both the delta-tubulin (*LOC51174*) and kinesin family member 5B (*KIF5B*) genes were also downregulated. These two genes play a role in spindle formation. In addition, genes related to transcription such as general transcription factor IIH, polypeptide 1 (*GTF2H1*), polymerase II RNA subunit C (*POLR2C*), DNA polymerase delta subunit 3 (*POLD3*), zinc finger protein 26 (*ZNF26*), and the zinc finger protein 317 (*ZNF317*) were all upregulated.

DISCUSSION

Coeliac disease is a chronic inflammatory disorder triggered by dietary gluten that shows major manifestations in the small intestine although other organ systems may also be involved. Coeliac disease is an HLA associated disease and over the past years our understanding of the role of both HLA-DQ2 and -DQ8 molecules and gluten polypeptides has been enhanced tremendously.[2] Nevertheless, a detailed understanding of the molecular events that take place in the intestine of genetically susceptible individuals is lacking, as well as insight into the order in which the events occur and other genetic factors that are known to be important determinants of the disease. To unravel novel aspects of the pathogenesis of coeliac disease, cDNA microarray technology was used to monitor expression of thousands of genes simultaneously. The results from this study confirm earlier studies that coeliac disease is a Th1 mediated disease,[4 21] as evidenced by upregulation of the IL-2 pathway. Interestingly, novel genes are being identified that have not been proposed previously as being important determinants of the pathogenesis of coeliac disease. These genes yield new insights into the molecular processes underlying the flattened mucosa.

Prolyl endopeptidase

The most notable gene differentially expressed in a comparison of coeliac disease patients on a gluten containing diet versus those on a gluten free diet, is prolyl endopeptidase (*PREP*). The *PREP* gene encodes a cytosolic prolyl endopeptidase that efficiently hydrolyses proline rich fragments such as gliadin. *PREP* is upregulated in coeliac disease patients on a gluten containing diet compared with patients adhering to a gluten free diet. Furthermore, its levels of expression stay elevated in controls. Interestingly, this gene was also shown to be upregulated in seven control samples not on a gluten free diet compared with four biopsies from coeliac disease patients on a gluten free diet ($p<0.003$) and, to a lesser extent, when compared with the four refractory coeliac disease samples on a gluten free diet ($p<0.025$). This observation indicates that it is the presence of gluten in the intestine which modulates expression of this gene. It has been hypothesised that prolyl endopeptidase cleaves gliadin in small fragments in the intestine.[22] We hypothesise that, in coeliac disease patients, the activity of this molecule may be impaired and therefore it does not cut gliadin into sufficiently small fragments to create motifs long enough to be recognised by reactive T cells.[22] It has been proposed that a bacterial prolyl endopeptidase from *Flavobacterium memningoseticum* may be a target for future treatment of coeliac disease patients.[23] Further research should be done in order to elucidate the mechanisms controlling activation and expression of this protein in the intestines of normal and coeliac disease patients.

Molecular mechanisms controlling hyperplasia and villous atrophy in MIII biopsies

Enterocytes, one of the main components of the villous, arises from pluripotent stem cells located in the base of the intestinal crypts. When these pluripotent stem cells stop dividing, they complete their programme of differentiation and start migrating towards the apical part of the villous. Non-dividing epithelial cells located in the crypts of the small intestine that are ready to migrate and differentiate express the TM4SF4 gene.[24] In vitro studies have suggested that TM4SF4 plays a crucial role in regulation of proliferation and differentiation along the crypts by inhibiting proliferation of cells at the boundary of the crypt and villous. We observed an eightfold decrease in the level of expression of the TM4SF4 gene in MIII biopsies of coeliac disease patients versus M0 biopsies of controls, implying a block in early differentiation and failure to complete villous maintenance. Furthermore, increased expression of the *RBP4* gene and significant upregulation ($p<0.01$) of the Eph related receptor tyrosine kinase B3 (*EphB3*) gene in MIII versus M0 biopsies was observed. EphB3 has recently been shown to be tightly regulated by the Wnt cascade and it has been suggested that this gene may play a decisive role in controlling not only proliferation and differentiation but also migration and location of cells along the crypts.[25]

Given these results, we propose that stem cells of the small intestinal crypts start proliferating but do not receive a signal to start differentiation, leading to the formation of undifferentiated hyperplastic crypts and subsequently villous atrophy. This theory is further substantiated by results from the second experiment in which MIII lesions with or without exposure to gluten were compared. The presence of gluten leads to increased mitotic activity, suggesting many proliferating cells. As soon as gluten is withdrawn from the diet, mitotic activity diminishes, suggesting that cells stop proliferating and, presumably, start differentiating. Although the histology still shows a MIII lesion, this event might actually be the first step towards normalisation of the intestine. We further propose that the separate processes of crypt hyperplasia and villous atrophy are molecularly intertwined and progression through different histological stages may be regulated by genes implicated in the control of cell proliferation and differentiation, such as TM4SF4.

So far, the early pathogenic events leading to villous atrophy have been suggested to be direct effects of cytokines, such as IFN-γ,[21] or hypoxia.[3] In addition, using an ex vivo model of fetal gut, Salmela and colleagues recently proposed that increased levels of metalloproteinases are important molecules for tissue remodelling and mucosal degradation in inflammatory bowel disease and coeliac disease.[26] However, none of the 10 metalloproteinase family members, nor two of its inhibitors that were present on our slides, showed a significant change in expression level in MIII biopsies of coeliac disease patients (data not shown), implying that these molecules may not play a direct role in the tissue damage in vivo. Until now, all studies have proposed that villous atrophy is caused by *destruction* of enterocytes and, consequently, loss of structure of the villous. We propose that villous atrophy is not due to destruction of the villous but rather to the failure of crypt cells to differentiate into fully differentiated villous cells. As the cells seem to be capable of continuing proliferation, this might explain crypt hyperplasia.

We have successfully demonstrated the use of microarrays as a general approach to studying complex human diseases. Expression profiles of the selected biopsies propose known genes that participate in the pathogenesis of coeliac disease as well as new pathways that may play a potential role in disease initiation and progression. Our results suggest that

crypt hyperplasia may be explained by distortion in the ratio between cell proliferation and cell differentiation of cells composing the crypt-villi units. Further investigations of the identified genes are required and will hopefully advance our understanding of the molecular mechanisms underlying coeliac disease. Finally, the finding that prolyl endopeptidase is differentially expressed in coeliac disease patients may provide new leads for therapy.

ACKNOWLEDGEMENTS

The authors thank Dr Frits Koning and Darius Stepniak for helpful discussions, Daniel Chan for helping with the illustration, Justin Fransen for the real time RT-PCR, Jackie Senior for critically reading the manuscript, and the coeliac disease families for their interest and collaboration. This work was financially supported by the Dutch Digestive Disease Foundation (MLDS) grant WS00-13, the Netherlands Organisation for Medical and Health Research (NWO) grant 902-22-094, the KNAW Ter Meulen Fonds, and a Fulbright fellowship to CW.

Authors' affiliations

B Diosdado, M C Wapenaar, L Franke, C Wijmenga, Complex Genetics Group, Department of Biomedical Genetics, University Medical Centre, Utrecht, the Netherlands
K J Duran, Complex Genetics Group, and Genomics Laboratory, Department of Biomedical Genetics, University Medical Centre, Utrecht, the Netherlands
M J Goerres, Department of Gastroenterology and Hepatology, Rijnstate Hospital, Arnhem, the Netherlands
M Hadithi, C J J Mulder, Department of Gastroenterology and Hepatology, Rijnstate Hospital, Arnhem, the Netherlands, and Department of Gastroenterology, Free University Medical Centre, Amsterdam, the Netherlands
J B A Crusius, Laboratory of Gastrointestinal Immunogenetics, Free University Medical Centre, Amsterdam, the Netherlands
J W R Meijer, Department of Pathology, Rijnstate Hospital, Arnhem, the Netherlands
D J Duggan*, National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH, Bethesda, MD, USA
F C P Holstege, Genomics Laboratory, Department of Physiological Chemistry, University Medical Centre, Utrecht, the Netherlands

*Present address: TGen, Translational Genomics Research Institute, Phoenix, AZ, USA

REFERENCES

1 **Vader W**, Kooy Y, Van Veelen P, *et al.* The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. *Gastroenterology* 2002;**122**:1729–37.
2 **Sollid LM**. Molecular basis of celiac disease. *Annu Rev Immunol* 2000;**18**:53–81.
3 **Marsh MN**. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology* 1992;**102**:330–54.
4 **Sollid LM**. Coeliac disease: dissecting a complex inflammatory disorder. *Nat Rev Immunol* 2002;**2**:647–55.
5 **Schuppan D**. Current concepts of celiac disease pathogenesis. *Gastroenterology* 2000;**119**:234–42.
6 **Lockhart DJ**, Winzeler EA. Genomics, gene expression and DNA arrays. *Nature* 2000;**405**:827–36.
7 **Mulder C**, Rostami K, Marsh MN. When is a coeliac a coeliac? *Gut* 1998;**42**:594.
8 **Mulder CJ**, Wahab PJ, Moshaver B, *et al.* Refractory coeliac disease: a window between coeliac disease and enteropathy associated T cell lymphoma. *Scand J Gastroenterol Suppl* 2000;**232**:32–7.
9 **Van de Peppel JPK**, van Bakel H, Radonjic M, *et al.* Monitoring global mRNA changes with externally controlled microarray experiments. *EMBO* 2002;**4**:387–93.
10 **Brazma A**, Hingamp P, Quackenbush J, *et al.* Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat Genet* 2001;**29**:365–71.
11 **Chen Y**, Kamat V, Dougherty ER, *et al.* Ratio statistics of gene expression levels and applications to microarray data analysis. *Bioinformatics* 2002;**18**:1207–15.
12 **Chen Y**, Kamat V, Dougherty ER, *et al.* Ratio statistics of gene expression levels and applications to microarray data analysis. *Bioinformatics* 2002;**18**:1207–15.
13 **Livak KJ**, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)). *Methods Method* 2001;**25**:402–8.
14 **Nilsen EM**, Jahnsen FL, Lundin KE, *et al.* Gluten induces an intestinal cytokine response strongly dominated by interferon gamma in patients with celiac disease. *Gastroenterology* 1998;**115**:551–63.
15 **Nicoletti A**, Caligiuri G, Tornberg I, *et al.* The macrophage scavenger receptor type A directs modified proteins to antigen presentation. *Eur J Immunol* 1999;**29**:512–21.
16 **Yokota T**, Ehlin-Henriksson B, Hansson GK. Scavenger receptors mediate adhesion of activated B lymphocytes. *Exp Cell Res* 1998;**239**:16–22.
17 **Choudhury A**, Mukherjee P, Basu SK, *et al.* Disruption of T cell tolerance to self-immunoglobulin causes polyclonal B cell stimulation followed by inactivation of responding autoreactive T cells. *J Immunol* 2000;**164**:1713–21.
18 **Abraham R**, Choudhury A, Basu SK, *et al.* Disruption of T cell tolerance by directing a self antigen to macrophage-specific scavenger receptors. *J Immunol* 1997;**158**:4029–35.
19 **Angelini G**, Gardella S, Ardy M, *et al.* Antigen-presenting dendritic cells provide the reducing extracellular microenvironment required for T lymphocyte activation. *Proc Natl Acad Sci U S A* 2002;**99**:1491–6.
20 **Fahrer AM**, Konigshofer Y, Kerr EM, *et al.* Attributes of gammadelta intraepithelial lymphocytes as suggested by their transcriptional profile. *Proc Natl Acad Sci U S A* 2001;**98**:10261–6.
21 **Nilsen EM**, Lundin KE, Krajci P, *et al.* Gluten specific, HLA-DQ restricted T cells from coeliac mucosa produce cytokines with Th1 or Th0 profile dominated by interferon gamma. *Gut* 1995;**37**:766–76.
22 **Hausch F**, Shan L, Santiago NA, *et al.* Intestinal digestive resistance of immunodominant gliadin peptides. *Am J Physiol Gastrointest Liver Physiol* 2002;**283**:G996–1003.
23 **Shan L**, Molberg O, Parrot I, *et al.* Structural basis for gluten intolerance in celiac sprue. *Science* 2002;**297**:2275–9.
24 **Wice BM**, Gordon JI. A tetraspan membrane glycoprotein produced in the human intestinal epithelium and liver that can regulate cell density-dependent proliferation. *J Biol Chem* 1995;**270**:21907–18.
25 **Batlle E**, Henderson JT, Beghtel H, *et al.* Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/EphrinB. *Cell* 2002;**111**:251–63.
26 **Salmela MT**, MacDonald TT, Black D, *et al.* Upregulation of matrix metalloproteinases in a model of T cell mediated tissue injury in the gut: analysis by gene array and in situ hybridisation. *Gut* 2002;**51**:540–7.

Chapter 5

Arrested enterocyte differentiation in coeliac disease causes clinical pleiotropism and IBD-like detoxification impairment

MC Wapenaar, B Diosdado, H van Bakel,
E van Oort, CJJ Mulder, and C Wijmenga

(Submitted for publication)

ABSTRACT

Background & Aims: Coeliac disease (CD) is a gluten-induced malabsorption syndrome, marked by villous atrophy and a concomitant reduction of the absorptive surface. We aimed to determine whether events other than surface loss contributed to the nutrient deficiency. ***Methods***: Microarray expression profiles were made from the duodenal mucosa of untreated CD patients (n=13), treated patients in various stages of remission (n=31), and non-CD controls (n=21). Genes up-regulated during tissue recovery were identified by combined ANOVA analysis and correlation with the profile of the brushborder gene *DPP4*. ***Results***: After selection and filtering, 242 up-regulated genes were identified, most of which were characterized by a profile of gradual and progressive increase, synchronous with tissue normalization. Transcriptional increase progressed beyond the point where villi had already recovered. Diverse biological processes were involved, including the absorption and metabolism of sugars, peptides, fat, sterols, iron, and water. Loss of gene activity was associated with the pleiotropic clinical features of CD. We also observed a remarkable loss of detoxification capacity in CD, similar to that reported for inflammatory bowel disease (IBD). All our data pointed towards arrested differentiation of enterocytes in CD, supported by the transcriptional profiling and immunohistochemistry of the differentiation-promoting gene *TM4SF4*. ***Conclusions***: CD is not only characterized by malabsorption due to villous atrophy but also by the consequences of arrested enterocyte differentiation. This impaired differentiation is only partly recovered after the villi have reappeared. This should be taken into consideration when dealing with gluten-sensitive patients who do not display the classic picture of atrophic villi, but who could still be affected by partial malabsorption.

INTRODUCTION

The response of coeliac disease (CD) patients to the cereal protein gluten is characterized by an intestinal inflammation that initiates a sequence of events in the mucosa, ultimately leading to a severe functional impairment. The successive steps in the mucosal remodeling involve the infiltration with lymphocytes, the development of hyperplastic crypts, and progressive atrophy of the villi[1]. This appears to be a tightly controlled reciprocal process, as atrophic patients on a gluten-free diet traverse this tissue transformation in reverse order, till the mucosa is completely normalized. This feature of CD makes it particularly attractive to study its pathogenesis, and more generally, to use it as a model for the regenerative processes in the intestine. Clinically, CD patients are characterized by the classic features of diarrhea, abdominal distention and discomfort, and malnutrition. However, long-term gluten exposure may lead to severe health-related problems that occur either in isolation or as a cluster and they may involve chronic fatigue,

growth retardation, anemia, osteopenia, infertility, spontaneous abortion, and neurological complaints[2].

So far it has been generally accepted that most, if not all, of these problems arise from macro- and micronutrient deficiency due to a loss in absorptive intestinal surface, directly related to the villous atrophy. In this study we demonstrate that, without negating the effect of atrophy, a hampered terminal differentiation of the enterocytes also leads to an incapacitated intestinal function that contributes to the pleiotropic clinical features of CD. By using global gene expression profiling we obtained direct insight into the diverse molecular processes perturbed. Interestingly, we observed a general impairment of detoxification in CD, a feature it has in common with inflammatory bowel disease (IBD)[3].

MATERIALS AND METHODS

Patient material

Clinical details on the patients with the Marsh classification of their duodenal biopsy samples are summarized in supplementary table 1. In total, 44 CD patients and 21 non-CD controls participated. Patients were diagnosed according to the UEGW criteria[4]. Thirteen of them were untreated at the time of diagnosis and their biopsies showed Marsh III lesions (villous atrophy, crypt hyperplasia, and lymphocytosis). The remaining 31 CD patients were being treated with a gluten-free diet and were in various stages of remission: MIII (n=3), MII (n=9), MI (n=8), and M0 (n=11). Approval for this study was obtained from the Medical Ethics Committee of the University Medical Center Utrecht.

RNA isolation and microarray hybridizations

Two to three duodenal biopsies were obtained from each individual by forced forceps endoscopy, and stored in liquid nitrogen. A small bowel resection was used as a common reference sample to normalize all microarray hybridizations. Frozen biopsies were homogenized and total RNA was isolated as previously described[5]. RNA from the reference and the biopsy samples were amplified. RNA amplification was performed with 1 µg of total RNA by in vitro transcription with T7 RNA-polymerase and 5-(3-aminoallyl)-UTP following the manufacture s instructions (Ambion Inc. TX, USA). Next, 1000 ng of each cRNA sample was used to incorporate either Cy3 or Cy5 label (Amersham Pharmacia Biotech, NJ, USA) as previously described[6]. Separate pools were made from Cy3- and Cy5-labeled reference cRNA. Samples used for hybridization had incorporation frequencies of labeled nucleotides typically between 1.5-2%. Dye-swap experiments were performed for all biopsy samples against the common reference (intestinal resection). In total, 130 hybridizations were carried out with microarray slides containing 21,521 70-mer oligonucleotides (Operon human version 2;

www.microarrays.med.uu.nl/microarrays/human2.phpsite). Samples were hybridized as dye-swap duplicates using 500 ng of either Cy3- or Cy5-labeled cRNA with the common reference labeled with the complementary dye. After overnight hybridization at 37°C, slides were washed manually and scanned with an Agilent G256AA DNA microarray scanner (100% laser power, 30% photomultiplier tube). All microarray expression data and protocols were deposited in the MIAME compliance database ArrayExpress[7].

Data analysis

Imagene v.5.6 software (BioDiscovery Inc., CA, USA) was used to quantify microarray signals. Data normalization on a sub-grid basis was performed with marrayNorm R package v.1.1.3, and its variance was stabilized with the VSN R package v.1.3.2, as described previously[8]. Relative expression of each gene was calculated as the ratio of each individual biopsy sample and the common intestinal tissue reference. Differentially expressed genes among the various patient and control groups were identified by ANOVA analysis ($P < 0.05$), using the MAANOVA R package v.0.95-3[9]. A fixed-effects model was used that takes into account array and dye effects, and only genes with a false-discovery rate (FDR) adjusted tabulated p-value of $q<0.05$ were selected.

The majority of the MAANOVA-selected genes, including the classic brushborder marker *DPP4*, displayed a consistent incremental increase in expression, coinciding with the Marsh recovery sequence (MIII-UT-MIII–MII–MI–M0) (where UT stand for untreated). We took advantage of this feature by selecting additional genes with an expression profile highly correlated ($r^2 > 0.95$) with that of *DPP4*, and filtered out those genes for which the ratio MIII-UT/M0 < 1.25 and the ratio MIII-UT/ NC < 1.25 (Where NC stands for non-CD control). Standard correlation selection and filtering was performed with GeneSpring v.6.1 (Silicon Genetics, Redwood City, CA, USA). Expression data of individuals in the same Marsh class were averaged and normalized against the average signal of the MIII-UT group. Functional annotations of the differentially expressed genes were derived from a PubMed literature study and Gene Ontology (www.geneontology.org).

Immunohistochemistry

Paraffin embedded duodenal biopsy sections (5-6 μm) were used from CD patients (MIII) and healthy controls. Antigen retrieval was performed by 15 min. incubation with Pepsine in a glycine/HCl buffer, pH 2 (2.5 10E6 units/L, 0.1 M glycine). Incubation of anti-TM4SF4 (Wice, 1995) (1:4000) was performed overnight at 4°C in phosphate buffered saline (PBS) with 1% bovine serum albumin (BSA). Powervision Rabbit-HRP (dilution 1:10000, DPVR-55HRP) was incubated for 30 min. Sections were rinsed in PBS and incubated for 3 min. in phosphate

citrate buffer pH 5.8, followed by a 10 min. incubation with 2mM DAB (3-3'diaminobenzidinetetrahydrocloride) and 0.03 % H_2O_2 in phosphate citrate buffer pH 5.8 for color development.

RESULTS

Differentially expressed genes and their molecular processes

We used MAANOVA statistics ($P < 0.05$) and expression profile similarity to the brushborder gene *DPP4* ($r^2 > 0.95$; MIII-UT/M0 > 1.25 or MIII-UT/NC >

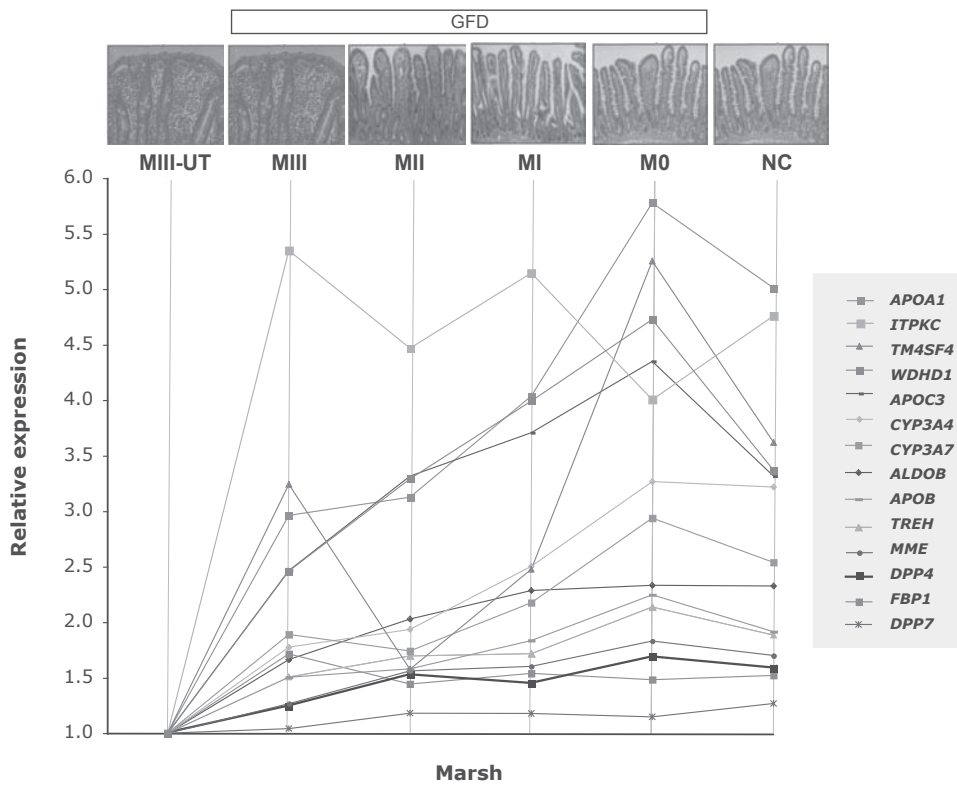


Figure 1. Representative expression profiles of 14 genes that show an incremental increase during the remission process of CD patients on a gluten-free diet. The mucosal histology of the Marsh recovery sequence is indicated in the upper part of the figure, and indicates villous atrophy (MIII), crypt hyperplasia (MII), lymphocytosis (MI), and complete remission (M0). Averaged relative expression values, after normalization to the untreated CD Marsh III patients (MIII-UT), are shown in the graph. The profile of brush border marker *DPP4* (black squares) was used to identify more genes with a similar profile. Details on the genes depicted can be found in table 1.

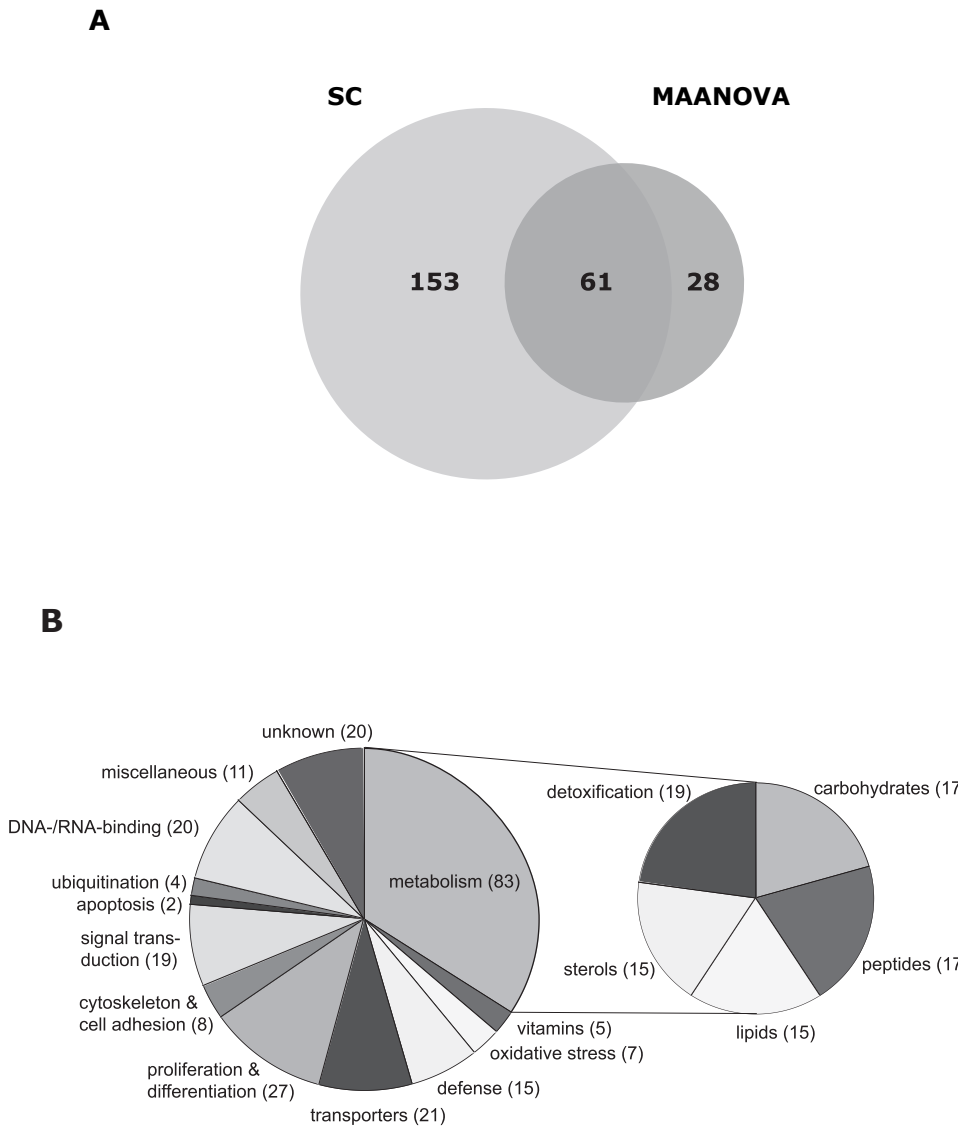


Figure 2. Numbers and functions of genes up-regulated during the remission process in CD. Venn diagram (**A**), indicating the number of genes selected using either MAANOVA statistics or Standard Correlation (SC), and the intersection of both methods. Pie charts (**B**), illustrating the type of processes involved and, in brackets, the number of genes they contain. The 'metabolism' genes are further categorized according to the five main functional processes relevant to this study.

1.25) to select a total of 242 genes that were up-regulated in the recovering CD duodenal mucosa. The expression profiles of 14 of these genes (including *DPP4*), representing various molecular processes, are shown in figure 1. The vast majority of the differentially expressed genes displayed this profile of a gradual and incremental increase during a remission that is marked by clearly defined histological transitions. Of the 242 up-regulated genes, 89 were identified by MAANOVA analysis, and 214 by similarity to the *DPP4* profile (figure 2A).

The 61 genes selected by both analysis methods were typically characterized by a smoothly increasing profile and a marked differential expression (table 1). Although twice as many genes were identified with standard correlation compared to MAANOVA, there was no apparent difference in the spectrum of molecular processes that they represented (table 1). This is further underscored by the fact that 70% of the MAANOVA-selected genes were also recognized by *DPP4*-correlation. In figure 2B we have grouped the up-regulated genes according to the molecular and cellular processes they represent. The largest group (83 genes) is that which we have lumped under the nominator 'metabolism', and which includes five main functions involved in nutrient absorption and processing relevant to the clinical picture of CD. Other major activated functions include proliferation and differentiation, transport, gene regulation, and signal transduction, and these possibly point more to the morphogenetic aspects of the mucosal transformation. More detailed information on the annotation and expression of the genes is given in table 1. It was remarkable that of the 242 genes, 59 (24%) could be attributed to epithelial cells, and 30 (12%) more specifically to the brushborder. This was even more evident in the five major nutrient absorption/processing groups: out of 83 genes, 37 (45%) were expressed in epithelial cells, and 16 (19%) in the brushborder (table 1). This is probably an underestimation, due to a lack of functional annotation, but it clearly demonstrates that, despite using whole mucosal biopsies, valuable information can be obtained on the transcriptional activity of enterocytes, even those confined to specific apical functions.

Impaired functions and clinical pleiotropism

The wide range of impaired molecular processes in CD, as judged from the expression profiles, have compelling links to the wide-range nutrient malabsorption and diverse clinical features observed in the patients. In table 2 we have attempted to attach functional processes and specific genes to known clinical symptoms, although we are well aware that a one-to-one comparison will have limitations due to the concerted involvement of a multiplicity of genes within a myriad network. It is important to realize that because of this, entire integrated functional processes could be undermined by the gluten-evoked response.

Table 1. Recovery of functional processes with increased gene expression in CD patients, in remission on a gluten-free diet

Gene symbol	Expression	Chromosome	Statistics		Relative expression in CD patients and controls					
			MA	SC	MIII-UT	MIII	MII	MI	M0	NC
Carbohydrate processing										
ALDOB	EPTH	9q21.3-q22.2	+	+	1.00	1.66	2.03	2.28	2.33	2.33
ALDOC	EPTH	17cen-q12	-	+	1.00	1.06	1.18	1.13	1.39	1.27
FBP1	EPTH	9q22.3	+	-	1.00	1.71	1.44	1.54	1.48	1.52
GALNT1	EPTH	18q12.1	-	+	1.00	0.99	1.25	1.17	1.38	1.24
GALT	EPTH	9p13	-	+	1.00	1.03	1.19	1.10	1.23	1.24
IDH2		15q26.1	-	+	1.00	1.13	1.52	1.36	1.37	1.49
ME1	EPTH	6q12	-	+	1.00	1.22	1.35	1.18	1.21	1.27
PCK1	BBM	20q13.31	+	+	1.00	1.30	1.59	1.51	1.49	1.53
PDK4		7q21.3	+	+	1.00	1.36	1.66	1.48	1.59	1.54
PGM1		1p31	-	+	1.00	1.20	1.39	1.25	1.34	1.25
SLC2A2	BBM	3q26.1-q26.2	+	+	1.00	1.38	1.51	1.53	1.65	1.48
SLC2A5	BBM	1p36.2	+	+	1.00	1.51	1.44	1.42	1.74	1.31
SLC2A12	EPTH	6q23.2	-	+	1.00	1.16	1.25	1.25	1.33	1.27
SLC5A1	BBM	22q13.1- q12.3	+	+	1.00	1.24	1.35	1.36	1.58	1.40
SMPDL3A		6q22.31	-	+	1.00	1.18	1.19	1.13	1.37	1.24
SORD	EPTH	15q15.3	-	+	1.00	1.18	1.30	1.29	1.50	1.43
TREH	BBM	11q23.3	+	+	1.00	1.51	1.70	1.71	2.14	1.88
Peptide and amino acid processing										
ACY1	EPTH	3p21.1	-	+	1.00	1.15	1.42	1.24	1.36	1.24
ALDH18A1	EPTH	10q24.3	-	+	1.00	1.18	1.29	1.19	1.39	1.24
ANPEP	BBM	15q25-q26	+	+	1.00	1.19	1.41	1.50	1.61	1.30
CAPN13		2p22-p21	+	-	1.00	2.12	1.60	1.64	1.64	1.62
CTSE	BBM	1q31	-	+	1.00	1.39	1.43	1.42	1.56	1.42
DPP4	BBM	2q24.3	+	+	1.00	1.24	1.53	1.45	1.69	1.59
DPP7		9q34.3	-	+	1.00	1.04	1.18	1.18	1.15	1.27
GATM	INTN	15q21.1	+	+	1.00	1.18	1.54	1.47	1.83	1.56
GLS	EPTH	2q32-q34	-	+	1.00	1.06	1.25	1.16	1.37	1.32
MEP1A	BBM	6p12-p11	+	+	1.00	1.32	1.51	1.44	1.71	1.60
MME	BBM	3q25.1-q25.2	+	+	1.00	1.27	1.56	1.60	1.83	1.70
OAT	EPTH	10q26	+	+	1.00	1.32	1.54	1.25	1.58	1.47
P4HA2		5q31	-	+	1.00	1.57	1.54	1.44	1.56	1.85
PEPD	EPTH	19q12-q13.2	-	+	1.00	1.25	1.47	1.36	1.48	1.35
SLC7A7	EPTH	14q11.2	-	+	1.00	1.04	1.23	1.17	1.34	1.24
SLC15A1	BBM	13q33-q34	+	+	1.00	1.18	1.45	1.48	1.89	1.40
VPS35	EPTH	16q12	+	+	1.00	1.16	1.34	1.34	1.50	1.55

Gene symbol	Expression	Chromosome	Statistics		Relative expression in CD patients and controls					
			MA	SC	MIII-UT	MIII	MII	MI	M0	NC
Lipid processing										
ACADM		1p31	-	+	1.00	1.25	1.43	1.38	1.49	1.38
ACAT1	EPTH	11q22.3-q23.1	+	+	1.00	1.09	1.46	1.61	1.72	1.78
ACSS2	INTN	20q11.22	+	+	1.00	1.70	1.49	1.57	1.69	1.51
APOA1	EPTH	11q23-q24	+	+	1.00	2.96	3.12	4.04	5.78	5.01
APOB	BBM	2p24-p23	+	+	1.00	1.50	1.58	1.84	2.25	1.91
APOC2	EPTH	19q13.2	+	+	1.00	1.53	1.87	1.73	2.19	2.06
APOC3		11q23.1-q23.2	-	+	1.00	2.46	3.31	3.71	4.35	3.31
CD36	BBM	7q11.2	+	+	1.00	1.25	1.56	1.63	1.93	1.78
ECH1		19q13.1	-	+	1.00	1.08	1.25	1.18	1.22	1.32
FABP2	EPTH	4q28-q31	-	+	1.00	1.13	1.30	1.35	1.41	1.29
HACL1		3p24.3	-	+	1.00	1.24	1.44	1.44	1.51	1.38
HADHB		2p23	-	+	1.00	1.08	1.34	1.31	1.39	1.38
HMGCS2		1p13-p12	-	+	1.00	1.43	1.83	1.91	2.16	2.13
LIPA		10q23.2-q23.3	+	+	1.00	1.45	1.61	1.52	1.70	1.64
LIPF	EPTH	10q23.31	+	-	1.00	1.03	2.28	1.30	1.49	1.12
Sterol/steroid/endocrine processing										
ABCG5	INTN	2p21	-	+	1.00	1.15	1.25	1.43	1.55	1.54
ADM	EPTH	11p15.4	-	+	1.00	1.24	1.29	1.35	1.35	1.41
ANGPTL4	INTN	19p13.3	-	+	1.00	1.26	1.35	1.15	1.25	1.29
CTSH	EE	15q24-q25	-	+	1.00	1.25	1.26	1.22	1.34	1.35
DHRS8	EPTH	4q22.1	+	+	1.00	1.41	1.31	1.36	1.60	1.51
HPGD	EPTH	4q34-q35	-	+	1.00	1.19	1.18	1.22	1.43	1.28
HSD17B2	EPTH	16q24.1-q24.2	-	+	1.00	1.40	1.43	1.38	1.44	1.45
NCOA4	EPTH	10q11.2	-	+	1.00	1.23	1.25	1.35	1.33	1.35
NELF		9q34.3	-	+	1.00	1.05	1.21	1.23	1.39	1.24
PGRMC1	EPTH	Xq22-q24	-	+	1.00	0.97	1.22	1.20	1.30	1.32
PGRMC2		4q26	-	+	1.00	1.17	1.11	1.27	1.45	1.30
SCG3	EE	15q21	-	+	1.00	1.29	1.49	1.35	1.26	1.46
SCP2	EPTH	1p32	-	+	1.00	1.07	1.29	1.18	1.31	1.27
SC5DL		11q23.3	-	+	1.00	1.18	1.14	1.26	1.25	1.31
TRHDE		12q15-q21	+	+	1.00	1.35	1.40	1.47	1.68	1.47
Vitamins										
ADH4	EPTH	4q22	-	+	1.00	1.35	1.50	1.46	1.37	1.48
DHRS3		1p36.1	-	+	1.00	1.38	1.28	1.24	1.37	1.37
GC		4q12-q13	-	+	1.00	1.08	1.30	1.35	1.27	1.27
RDH5	EPTH	12q13-q14	-	+	1.00	1.05	1.19	1.23	1.44	1.34
SLC23A1	BBM	5q31.2-q31.3	-	+	1.00	1.33	1.40	1.29	1.42	1.42
Detoxification										
ABCC2	BBM	10q24	+	+	1.00	1.33	1.34	1.38	1.70	1.38
ABCG2	BBM	4q22	-	+	1.00	1.39	1.44	1.27	1.45	1.25

Gene symbol	Expression	Chromosome	Statistics		Relative expression in CD patients and controls					
			MA	SC	MIII-UT	MIII	MII	MI	M0	NC
CYP2C9	EPTH	10q24	+	+	1.00	1.27	1.37	1.45	1.64	1.38
CYP2C18	EPTH	10q24	-	+	1.00	1.19	1.21	1.21	1.24	1.27
*CYP3A4**	EPTH	7q21.1	+	+	1.00	1.77	1.93	2.50	3.26	3.22
*CYP3A4**	EPTH	7q21.1	+	+	1.00	2.14	1.96	2.45	3.29	3.15
CYP3A7	EPTH	7q21-q22.1	+	+	1.00	1.89	1.74	2.17	2.94	2.54
DHDH	INTN	19q13.3	+	+	1.00	1.51	1.66	1.40	1.81	1.43
GAL3ST1	EPTH	22q12.2	-	+	1.00	1.25	1.41	1.46	1.60	1.47
GBA3		4p15.31	-	+	1.00	1.02	1.45	1.31	1.38	1.32
GGT1	BBM	22q11.23	-	+	1.00	1.16	1.48	1.17	1.41	1.32
GGTLA4		20p11.1	-	+	1.00	1.06	1.17	1.15	1.38	1.31
GSTA4	EPTH	6p12.1	-	+	1.00	1.14	1.19	1.07	1.21	1.24
FLJ16331	EPTH	11q13.1	+	-	1.00	1.23	1.10	1.45	1.64	1.50
SULT2A1	INTN	19q13.3	-	+	1.00	1.16	1.31	1.24	1.35	1.25
UGT1A9	EPTH	2q37	-	+	1.00	1.37	1.43	1.33	1.45	1.33
UGT1A9	EPTH	2q37	+	+	1.00	1.41	1.53	1.40	1.56	1.42
UGT2A3	INTN	4q13.2	+	+	1.00	1.51	1.74	1.70	1.85	1.80
UGT2B7	INTN	4q13	+	+	1.00	1.13	2.06	2.69	2.92	2.93
Oxidative stress										
DHRS7		14q23.1	-	+	1.00	1.40	1.33	1.40	1.37	1.45
GPX3	INTN	5q23	-	+	1.00	1.26	1.26	1.21	1.25	1.29
MSRA	EPTH	8p23.1	-	+	1.00	1.22	1.26	1.32	1.23	1.37
NQO2		6pter-q12	-	+	1.00	1.24	1.43	1.20	1.39	1.35
PRDX3		10q25-q26	+	-	1.00	0.73	0.91	1.69	1.05	1.21
TXNRD2	EPTH	22q11.21	-	+	1.00	1.39	1.71	1.50	1.43	1.43
VNN1	BBM	6q23-q24	+	+	1.00	1.54	1.83	1.89	2.20	1.94
Defense										
CXCL14		5q31	+	+	1.00	1.37	1.51	1.64	1.61	1.62
CYP4F2	INTN	19pter-p13.11	+	+	1.00	1.61	1.62	1.65	2.13	1.68
F10	EPTH	13q34	+	+	1.00	1.30	1.47	1.77	1.84	1.98
FAM3C	INTN	7q31	-	+	1.00	1.18	1.23	1.28	1.29	1.24
IRAK4	INTN	12q12	-	+	1.00	1.12	1.28	1.30	1.21	1.26
LEAP-2	INTN	5q31.1	-	+	1.00	1.35	1.29	1.37	1.53	1.25
PAG1		8q21.13	+	+	1.00	1.20	1.42	1.41	1.39	1.70
PPM1A		14q23.1	-	+	1.00	1.07	1.16	1.14	1.27	1.24
PRAP1		10q26.3	+	+	1.00	1.63	1.65	1.81	2.05	1.87
PYY2	EPTH	17q11	+	+	1.00	1.46	1.35	1.59	1.75	1.60
RNF128		Xq22.3	-	+	1.00	1.50	1.39	1.41	1.40	1.50
SPINK1	INTN	5q32	-	+	1.00	1.34	1.18	1.40	1.47	1.47
TFPI	EPTH	2q32	+	-	1.00	0.92	1.29	1.24	1.64	1.38
TREM2		6p21.1	-	+	1.00	1.26	1.63	1.30	1.36	1.49
TNFRSF11A		18q22.1	-	+	1.00	1.34	1.47	1.39	1.39	1.35

Gene symbol	Expression	Chromosome	Statistics		Relative expression in CD patients and controls					
			MA	SC	MIII-UT	MIII	MII	MI	M0	NC
Transporters										
AP3B2		15q	-	+	1.00	1.16	1.36	1.24	1.26	1.26
AQP10	BBM	1q21.3	+	+	1.00	1.45	1.83	2.04	2.06	1.87
ATP9A		20q13.2	+	-	1.00	0.64	0.97	1.51	0.93	1.04
CYBRD1	BBM	2q31.1	+	+	1.00	1.44	1.88	2.14	2.16	2.29
GOSR1		17q11	-	+	1.00	1.28	1.27	1.24	1.26	1.34
HBB		11p15.5	-	+	1.00	1.59	1.53	1.65	1.59	1.70
HCP1	BBM	17q11.2	-	+	1.00	1.00	1.20	1.29	1.42	1.33
HEBP1		12p13.1	+	-	1.00	1.98	1.48	1.40	1.54	1.33
KCNK15		20q12-q13	+	-	1.00	1.07	1.30	1.62	1.49	1.70
MOSC2		1q41	-	+	1.00	1.02	1.15	1.13	1.24	1.25
PSCD4		22q12.3-q13.1	+	-	1.00	1.55	1.22	1.81	1.34	1.64
PLLP	BBM	16q13	-	+	1.00	1.14	1.17	1.23	1.29	1.24
SLC4A4	EPTH	4q21	-	+	1.00	1.25	1.30	1.43	1.37	1.45
SLC6A13		12p13.3	-	+	1.00	1.21	1.34	1.21	1.29	1.35
SLC13A2	BBM	17p13.2	+	-	1.00	1.01	1.08	1.62	1.65	1.22
SLC17A4		6p22-p21.3	-	+	1.00	1.14	1.19	1.21	1.41	1.24
SLC25A24		1p13.3	-	+	1.00	1.28	1.29	1.22	1.34	1.23
SNX3		6q21	-	+	1.00	2.32	2.17	2.17	2.50	2.90
SNX24		5q23.2	-	+	1.00	1.04	1.24	1.18	1.33	1.25
SRP19		5q21-q22	-	+	1.00	1.28	1.15	1.22	1.29	1.28
TMEM37		2q14.2	-	+	1.00	1.13	1.14	1.07	1.29	1.27
Proliferation/differentiation/remodeling										
ABL1		9q34.1	+	+	1.00	1.31	1.28	1.59	1.50	1.58
ANXA13	BBM	8q24.13	+	+	1.00	1.36	1.47	1.29	1.63	1.66
BTG3		21q21.1-q21.2	+	+	1.00	1.34	1.42	1.33	1.72	1.66
CAP2		6p22.3	+	-	1.00	0.65	1.02	1.44	0.89	1.08
EFNA1	EPTH	1q21-q22	+	-	1.00	0.98	1.27	1.62	1.44	1.44
ENPP2		8q24.1	-	+	1.00	1.19	1.24	1.23	1.31	1.39
FZD2	INTN	17q21.1	+	-	1.00	0.64	1.38	1.01	1.37	0.99
IGFBP3		7p13-p12	-	+	1.00	1.19	1.25	1.13	1.30	1.34
IGFBP4	EPTH	17q12-q21.1	-	+	1.00	1.14	1.14	1.26	1.53	1.33
IHH		2q33-q35	-	+	1.00	1.21	1.22	1.24	1.34	1.30
ING4		12p13.31	-	+	1.00	1.35	1.50	1.42	1.60	1.42
LAMB2	EPTH	3p21	-	+	1.00	1.04	1.29	1.27	1.59	1.40
MATN2		8q22	-	+	1.00	1.01	1.29	1.25	1.37	1.40
MUCDHL		11p15.5	-	+	1.00	1.01	1.26	1.39	1.52	1.37
MXRA5		Xp22.33	-	+	1.00	1.30	1.29	1.11	1.46	1.39
NELL2	EPTH	12q13.11-q13.12	+	+	1.00	1.04	1.34	1.43	1.80	1.77
OTP		5q13.3	-	+	1.00	1.16	1.28	1.23	1.24	1.26
PCLKC	BBM	5q35.2	+	+	1.00	1.35	1.57	1.80	1.96	1.74

Gene symbol	Expression	Chromosome	Statistics		Relative expression in CD patients and controls					
			MA	SC	MIII-UT	MIII	MII	MI	M0	NC
PLOD2		3q23-q24	+	+	1.00	1.32	1.75	1.56	1.69	1.66
PPP1CC		12q24.1-q24.2	-	+	1.00	1.15	1.10	1.12	1.25	1.24
PRRG2	EPTH	19q13.33	+	-	1.00	0.95	1.46	1.09	1.53	1.14
PTPRF	EPTH	1p34	+	+	1.00	1.16	1.21	1.42	1.51	1.53
RGN		Xp11.3	-	+	1.00	1.15	1.17	1.17	1.26	1.28
SRC	EPTH	20q12-q13	-	+	1.00	1.08	1.24	1.35	1.34	1.27
TM4SF4	BBM	3q25	+	-	1.00	3.24	1.57	2.48	5.25	3.62
TM4SF5		17p13.3	-	+	1.00	1.47	1.56	1.47	1.79	1.47
TRIM45		1p13.1	-	+	1.00	1.31	1.29	1.37	1.36	1.50
Cytoskeleton and cell adhesion										
ACTA2		10q23.3	-	+	1.00	1.05	1.12	1.22	1.26	1.30
ACTR1B		2q11.1-q11.2	+	+	1.00	1.67	1.63	1.71	2.07	1.80
ANK3	EPTH	10q21	+	+	1.00	1.12	1.37	1.25	1.56	1.45
DSG3	EPTH	18q12.1-q12.2	-	+	1.00	1.12	1.29	1.28	1.46	1.24
EPB49		8p21.1	-	+	1.00	1.16	1.21	1.30	1.19	1.25
FILIP1		6q14.1	-	+	1.00	1.17	1.17	1.13	1.22	1.27
NCKAP1		2q32	-	+	1.00	1.10	1.19	1.16	1.25	1.30
PDLIM7		5q35.3	-	+	1.00	1.10	1.34	1.24	1.53	1.29
Signal transduction										
ABRA		8q23.1	-	+	1.00	1.07	1.19	1.23	1.20	1.30
AKAP9		7q21-q22	-	+	1.00	1.12	1.18	1.22	1.29	1.26
APLP2		11q24	-	+	1.00	1.09	1.29	1.38	1.45	1.39
CAMK2N1	(BBM)	1p36.12	+	+	1.00	1.30	1.21	1.41	1.53	1.42
CERK		22q13.31	-	+	1.00	1.14	1.18	1.42	1.40	1.42
DAMS		5q31.1	-	+	1.00	1.31	1.33	1.37	1.34	1.33
GAPVD1		9q33.3	-	+	1.00	1.22	1.25	1.21	1.23	1.28
GIPC2		1p31.1	-	+	1.00	1.13	1.43	1.51	1.42	1.52
GNG12		1p31.3	+	-	1.00	2.04	1.42	1.38	1.46	1.32
GPA33	EPTH	1q24.1	-	+	1.00	1.28	1.20	1.27	1.47	1.28
GUCA2B	BBM	1p34-p33	+	+	1.00	1.30	1.34	1.99	1.90	1.81
ITPKC	BBM	19q13.1	-	+	1.00	5.35	4.47	5.14	4.01	4.76
PLEKHM1		17q21.31	-	+	1.00	1.16	1.23	1.23	1.23	1.24
PLXNB1		3p21.31	-	+	1.00	1.21	1.13	1.21	1.27	1.34
PTPN4		2q14.2	+	-	1.00	1.13	1.64	1.07	1.84	1.06
RAB17		2q37.3	-	+	1.00	1.19	1.19	1.22	1.39	1.29
RABL5		7q22.1	-	+	1.00	1.21	1.32	1.26	1.40	1.34
REEP6		19p13.3	+	+	1.00	1.11	1.38	1.33	1.58	1.47
TAS2R10		12p13	-	+	1.00	1.11	1.31	1.36	1.22	1.42
Apoptosis										
CIDEC		3p25.3	-	+	1.00	1.15	1.32	1.34	1.28	1.38
FAM3B		21q22.3	-	+	1.00	1.10	1.10	1.17	1.18	1.25

Gene symbol	Expression	Chromosome	Statistics		Relative expression in CD patients and controls					
			MA	SC	MIII-UT	MIII	MII	MI	M0	NC
Ubiquitination										
FBXL3		13q22	-	+	1.00	0.98	1.21	1.23	1.28	1.26
PSMA1		11p15.1	-	+	1.00	1.13	1.19	1.21	1.16	1.25
UBE4B		1p36.3	-	+	1.00	1.02	1.21	1.23	1.26	1.24
ZNRF3	INTN	22q12.1	+	-	1.00	0.70	0.90	0.91	1.03	1.32
DNA/RNA-binding										
FALZ		17q24.3	-	+	1.00	1.20	1.23	1.30	1.35	1.27
FAM12A	EPTH	14q11.2	+	+	1.00	1.73	3.71	3.62	5.19	2.44
HELZ		17q24.2	-	+	1.00	1.17	1.42	1.29	1.25	1.37
HMGB3		Xq28	-	+	1.00	1.27	1.32	1.23	1.42	1.48
KLF9		9q13	-	+	1.00	1.22	1.23	1.38	1.28	1.25
MAFK		7p22.3	+	-	1.00	2.05	1.21	1.13	1.37	1.34
MEF2B		19p12	+	-	1.00	1.52	1.56	1.41	1.68	1.64
MORC3		21q22.13	+	-	1.00	1.55	0.97	1.18	1.34	1.59
MPHOSPH10		2p13.3	-	+	1.00	1.27	1.32	1.35	1.28	1.26
NIN	EPTH	14q22.1	-	+	1.00	1.16	1.58	1.27	1.36	1.28
POU2F3		11q23.3	-	+	1.00	1.29	1.31	1.46	1.32	1.59
PSKH1		16q22.1	-	+	1.00	1.14	1.11	1.14	1.18	1.25
RFX1	EPTH	19p13.1	-	+	1.00	1.27	1.41	1.61	1.38	1.41
RPL4		15q22	+	+	1.00	1.94	1.89	2.16	2.48	2.36
SNAPC2		19p13.3-p13.2	-	+	1.00	1.18	1.21	1.15	1.18	1.24
STAU1		20q13.1	-	+	1.00	1.22	1.38	1.24	1.29	1.33
SUPT5H		19q13	+	-	1.00	0.59	1.34	1.11	1.46	1.10
WDHD1		14q22.3	+	+	1.00	2.46	3.29	3.99	4.72	3.37
ZNF655		7q22.1	-	+	1.00	1.06	1.19	1.23	1.28	1.27
ZNF580		19q13.42	-	+	1.00	1.09	1.39	1.16	1.29	1.26
Miscellaneous functions										
ALPI	BBM	2q37.1	+	+	1.00	1.71	1.46	1.63	1.91	1.57
CDADC1		13q14.2	-	+	1.00	1.19	1.19	1.08	1.27	1.25
CIB2		15q24	+	+	1.00	1.12	1.31	1.39	1.52	1.58
CSN3		4q21.1	-	+	1.00	1.17	1.53	1.14	1.45	1.47
LHPP		10q26.13	-	+	1.00	1.05	1.13	1.26	1.26	1.29
NT5E		6q14-q21	-	+	1.00	1.19	1.54	1.33	1.52	1.39
PEX12		17q12	-	+	1.00	1.14	1.37	1.25	1.31	1.25
PXMP2		12q24.33	-	+	1.00	1.30	1.41	1.38	1.44	1.56
QPRT		16p11.2	-	+	1.00	1.16	1.27	1.12	1.29	1.37
REN	EPTH	1q32	+	+	1.00	1.15	1.47	1.67	1.79	1.75
SHMT1		17p11.2	-	+	1.00	1.12	1.43	1.35	1.40	1.33
Unknown functions										
ALS2CR11		2q33.1	+	-	1.00	0.57	1.15	0.85	1.30	1.17
C5orf21		5q15	-	+	1.00	1.16	1.16	1.23	1.21	1.24

Gene symbol	Expression	Chromosome	Statistics		Relative expression in CD patients and controls					
			MA	SC	MIII-UT	MIII	MII	MI	M0	NC
C20orf28		20pter-q11.23	+	-	1.00	0.49	1.52	0.99	1.40	1.06
CCDC3		10p13	+	-	1.00	0.88	1.18	1.22	1.39	1.43
COBL		7p12.1	-	+	1.00	1.32	1.24	1.27	1.45	1.27
FAM76A		1p35.3	-	+	1.00	1.48	1.62	1.47	2.24	1.40
FLJ23447		19p13.12	-	+	1.00	1.19	1.24	1.40	1.46	1.44
KIAA1328		18q12.2	-	+	1.00	1.12	1.10	1.15	1.22	1.26
LOC91893		11q23.2	-	+	1.00	1.43	1.43	1.56	1.55	1.42
LOC284422		19p13.3	-	+	1.00	1.26	1.35	1.41	1.47	1.38
LRRC45		17q25.3	-	+	1.00	1.40	1.71	1.79	1.60	1.84
MAWBP		10pter-q25.3	+	+	1.00	1.04	1.27	1.46	1.50	1.52
OAF		11q23.3	-	+	1.00	1.23	1.30	1.17	1.23	1.24
PDZK7		10q24.31	+	-	1.00	0.47	0.97	0.93	1.00	0.91
PRR8		7q36.3	-	+	1.00	1.09	1.17	1.18	1.23	1.25
TMEM98		17q11.2	-	+	1.00	1.17	1.24	1.19	1.32	1.27
TMEM133		11q22.1	-	+	1.00	1.31	1.51	1.44	1.46	1.51
TMEM149		19q13.12	-	+	1.00	1.28	1.50	1.28	1.42	1.31
TTC5		14q11.2	-	+	1.00	1.37	1.34	1.33	1.28	1.39
TTC13		1q42.2	+	-	1.00	0.68	0.95	1.50	0.91	1.19

*) Represents two different oligonucleotide sequences on the array corresponding to the same gene. Abbreviations (in order of appearance): MA, microarray ANOVA; SC, standard correlation; MIII-UT, Marsh stage III untreated; MIII – M0, Marsh stage III – 0 in remission on a gluten-free diet and sequentially recovered from villous atrophy (MII), crypt hyperplasia (MI), and lymphocytosis (M0); NC, non-CD control; EPTH, epithelial cell; BBM, brushborder membrane; INTN, intestine; EE, enteroendocrine cell.

Through the identification of *FBP1* and *PCK1* expression in enterocytes, other than just liver and kidney, it became apparent that the intestine plays an important role in gluconeogenesis[10, 11]. In addition, glucose and galactose uptake from the lumen is facilitated by *SLC5A1* (*SGLT1*) at the apical site, while fructose is transported by *SLC2A5 (GLUT5)*[12, 13]. Basolateral transport of all three sugars to the circulation is facilitated by *SLC2A2 (GLUT2)*, and mutation in this gene has been in reported in Fanconi-Bickel syndrome (Wright, 2003)[13]. Our expression data indicate that both the gluconeogenesis and the sugar transport are impaired, and may contribute to the energy depletion in CD patients. Sugar malabsorption is also indicated by the elevated caloric-value of patients' stool (CJJM, unpublished results). Impaired peptide processing, as indicated by loss of brushborder endopeptidase function of *DPP4, DPP7*, and *MME*, may together contribute to physical wasting and developmental problems. The intestine also plays a major role in fatty acid metabolism through the binding of fatty acids by *FABP1* and *FABP2*, and the formation chylomicrons and HDL particles for further transport to the circulation[14]. Chylomicrons are rich in *APOB, APOC2*, and particularly abundant in *APOA1, APOC3*[15], which form a gene cluster with *APOA4*[16]. It had been

Table 2. Impaired metabolic processes in coeliac disease due to arrested enterocyte differentiation and their possible relations to the diverse clinical features

Metabolic process impaired	Examples of genes	Related clinical features
Carbohydrate processing	*FBP1, PCK1, SLC2A2, SLC2A5, SLC5A1*	Carbohydrate malabsorption, fatigue
Peptide and amino acid processing	*ANPEP, DPP4, DPP7, MEP1A, MME*	Wasting, failure to thrive
Lipid processing	*APOA1, APOB, APOC2, APOC3, FABP2, LIPA*	Steatorrhea, fat malabsorption
Sterol/steroid/endocrine processing	*ABCG5, ADM, HPGD, HSD17B2, NCOA4*	Infertility, spontaneous abortion, alopecia
Detoxification	*ABCC2, CYP3A4, CYP3A7, UGT2B7*	Adenocarcinoma
Iron uptake	*HCP1, HEBP1, CYBRD1*	Anemia
Water absorption	*AQP10*	Diarrhea

suggested that sprue-like conditions could result in reduced apolipoprotein production[14], and our results show that this is definitely the case. One of the hallmarks of severe fat malabsorption in CD is the steatorrhea[2]. We also observed several down-regulated genes involved in the sterol and steroid metabolism and these may be related to the reported endocrine[17], and reproductive disorders associated with CD[18, 19]. Absorption of iron takes place in haem-form from animal food, requiring transport by *HCP1*[20], and free iron, usually Fe^{3+} that needs to be reduced to Fe^{2+} by *CYBRD1*[21]. Down-regulation of both these genes in CD could be the major cause for the frequently observed anemia. Decreased expression of *AQP10*, located at the apical site of the enterocyte and involved in water uptake[22], could pose a risk for dehydration or diarrhea to CD patients. The epithelial cells of the intestine are exposed to xenobiotics and endobiotics and a versatile detoxification system preserves the integrity of enterocytes, and is the first line of protection for the individual. Detoxification in the gut is facilitated by phase I genes *(CYP3A4, CYP3A7, CYP2C9)*, phase II genes *(UGT1A9, UGT2A3, GSTA4, SULT2A1)*, and transporters *(ABCC2, ABCG2)*[23, 24]. Impairment of this system in CD, as indicated by our results, is not known to be associated with for example a specific drug-sensitivity, probably due to detoxification by the liver. However, the chronic exposure of enterocytes to enterotoxins could contribute to the risk of developing CD-associated adenocarcinoma[25].

Arrested enterocyte differentiation

The concerted down-regulation of genes involved in the diverse nutrient absorption and transport processes points toward a loss in enterocyte functionality that is compatible with a prematurely terminated differentiation program. This is supported by the strongly reduced expression of the differentiation marker *TM4SF4* in CD, both as transcript (figure 1, and table 1), and as brushborder protein (figure

3). *TM4SF4* is known for its proliferation-inhibiting and differentiation-promoting effect[26].

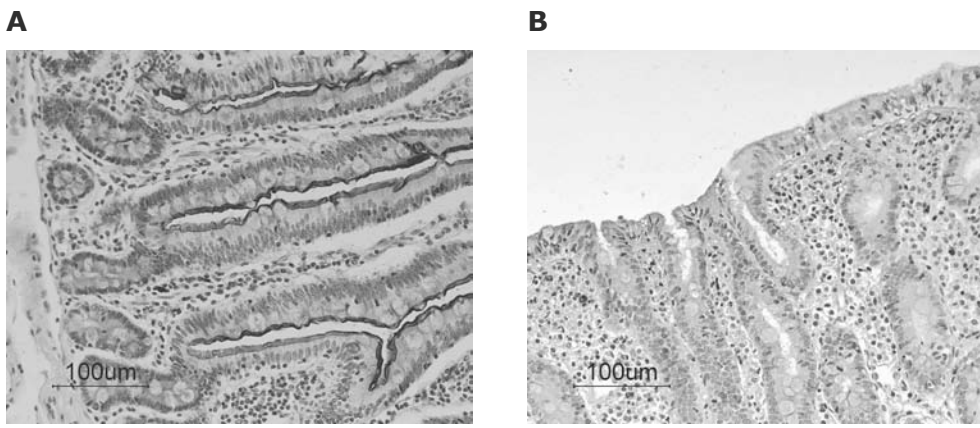


Figure 3. Immunohistochemistry on duodenal biopsy sections using anti-il-TMP antibodies against the *TM4SF4* gene product. TM4SF4 protein acts as a villus differentiation marker which is clearly present in the enterocyte brushborder of a control (**A**), but absent in the flat mucosa of a CD MIII patient (**B**).

DISCUSSION

We have presented a global gene expression profile of the duodenal mucosa in CD during recovery on a gluten-free diet. Since the tissue transformation in CD is reversible by withdrawing gluten from the diet, we may assume that this is the same for the gene transcription patterns. By taking advantage of samples from sequential steps in the pathology, as we did, it is possible to identify more differentially expressed genes than in conventional patient-control comparisons[27]. This is because a consistent and gradual change of the profile provides an additional selection parameter. Earlier we applied the same approach in a parallel study on the immune response in CD, using the same samples, but then concentrated on genes that diminished their expression during remission (Diosdado, submitted). An important aspect in this approach is that multiple patient samples per stage are used, preferably 10-20, as we have recently demonstrated (Wapenaar, submitted).

With the current study we make a compelling case for arrested enterocyte differentiation as part of the CD pathology. The resulting malabsorption effect adds to that created by the reduced absorptive surface due to the villous atrophy. That the observed differential expression was not simply an effect of the villous atrophy is evident from the continued increase in the gene profile long after the villi had reappeared in the stages MII – M0. Our observation that enterocytes have not completed terminal differentiation in the MII - M0 range has serious implications

for the medical counseling of atypical CD patients who do not reach the full-blown MIII stage[28]. Our data indicate that these patients still suffer, though less extensively, from an impaired intestinal function and that they should benefit from a gluten-free diet. The arrested enterocyte differentiation, affecting multiple cellular functions, is convincingly associated with the pleiotropic clinical features of the CD patients.

It is noteworthy that the loss of detoxification has also been reported in IBD[3]. The concerted involvement of multiple pregnane X receptor (PXR) target genes was interpreted as a loss of expression of this transcription factor. The role of PXR was further underscored by its genetic association with IBD susceptibility[29]. A model was proposed in which the epithelial barrier defect in IBD pathology is caused by dysfunctional detoxification as a result of genetically impaired transcriptional regulators[24]. However, based on our observations in CD, a similar dysfunctional detoxification in IBD could result from arrested terminal differentiation, possibly evoked by bacterial stimuli like gluten does in CD.

The question remains why gluten exposure in CD leads to incomplete epithelial differentiation. One possible explanation is that the inflammatory response induces enhanced stem cell proliferation[30], thereby accelerating the migration along the crypt-villus axis of the developing enterocytes, with insufficient time to complete their differentiation program. Alternatively, arrested differentiation, or partial loss of differentiation, may be part of an evolutionary program developed and carefully orchestrated to temporarily trade functionality to escape from attacks by parasites and pathogens. The unwanted and chronic activation of such a program by the gluten-evoked inflammation might have become an evolutionary faux pass with the relatively recent introduction of wheat-based food products in the human diet.

ACKNOWLEDGEMENTS

We extend our gratitude to the patients for their participation in the study. We would further like to thank Dr. J.W.R. Meijer (Rijnstate Hospital, Arnhem) for Marsh typing, Eric Strengman and Jan Geertzema for the immunohistochemistry work, and Jackie Senior for editing the manuscript. This work was financially supported by the Dutch Digestive Disease Foundation (MLDS, grant WS00-13), the Netherlands Organization for Scientific Research (NWO, grant 902-22-094, and VICI grant 918.66.620), and the Celiac Disease Consortium, an Innovative Cluster approved by the Netherlands Genomics Initiative and partially funded by the Dutch Government (BSIK03009).

REFERENCES

1. Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('coeliac sprue'). Gastroenterology 1992;102:330-54.
2. Green PH, Jabri B. Coeliac disease. Lancet 2003;362:383-91.
3. Langmann T, Moehle C, Mauerer R, et al. Loss of detoxification in inflammatory bowel disease: dysregulation of pregnane X receptor target genes. Gastroenterology 2004;127:26-40.
4. Report of a working group of the United European Gastroenterology Week in Amsterdam. When is a coeliac a coeliac? Eur J Gastroenterol Hepatol 2001;13:1123-28.
5. Wapenaar MC, van Belzen MJ, Fransen JH, et al. The interferon gamma gene in coeliac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility. J Autoimmun 2004;23:183-90.
6. Roepman P, Wessels LF, Kettelarij N, et al. An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. Nature genetics 2005;37:182-6.
7. Parkinson H, Sarkans U, Shojatalab M, et al. ArrayExpress--a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 2005;33:D553-5.
8. van Bakel H, Strengman E, Wijmenga C, et al. Gene expression profiling and phenotype analyses of S. cerevisiae in response to changing copper reveals six genes with new roles in copper and iron metabolism. Physiol Genomics 2005;22:356-67.
9. Wu HKK, Xiangqin C, Churchill GA. MAANOVA: A software package for the analysis of spotted cDNA microarray experiments. In: Parmigiani G, Garrett ES, Irizzary RA, Zeger SL, eds. The analysis of gene expression data: Methods and Software. New York: Springer-Verlag, 2003.
10. Yanez AJ, Nualart F, Droppelmann C, et al. Broad expression of fructose-1,6-bisphosphatase and phosphoenolpyruvate carboxykinase provide evidence for gluconeogenesis in human tissues other than liver and kidney. J Cell Physiol 2003;197:189-97.
11. Watford M. Is the small intestine a gluconeogenic organ. Nutr Rev 2005;63:356-60.
12. Mithieux G. The new functions of the gut in the control of glucose homeostasis. Curr Opin Clin Nutr Metab Care 2005;8:445-9.
13. Wright EM, Martin MG, Turk E. Intestinal absorption in health and disease--sugars. Best Pract Res Clin Gastroenterol 2003;17:943-56.
14. Riley JW, Glickman RM. Fat malabsorption--advances in our understanding. Am J Med 1979;67:980-8.
15. Zannis VI, Cole FS, Jackson CL, et al. Distribution of apolipoprotein A-I, C-II, C-III, and E mRNA in fetal human tissues. Time-dependent induction of apolipoprotein E mRNA by cultures of human monocyte-macrophages. Biochemistry 1985;24:4450-5.
16. Recalde D, Baroukh N, Viglietta C, et al. Human apoA-I/C-III/A-IV gene cluster transgenic rabbits: effects of a high-cholesterol diet. FEBS Lett 2004;572:294-8.
17. Collin P, Kaukinen K, Valimaki M, et al. Endocrinological disorders and coeliac disease. Endocr Rev 2002;23:464-83.

18. Meloni GF, Dessole S, Vargiu N, et al. The prevalence of coeliac disease in infertility. Hum Reprod 1999;14:2759-61.
19. Rostami K, Steegers EA, Wong WY, et al. Coeliac disease and reproductive disorders: a neglected association. Eur J Obstet Gynecol Reprod Biol 2001;96:146-9.
20. Shayeghi M, Latunde-Dada GO, Oakhill JS, et al. Identification of an intestinal heme transporter. Cell 2005;122:789-801.
21. Basu TK, Donaldson D. Intestinal absorption in health and disease: micronutrients. Best Pract Res Clin Gastroenterol 2003;17:957-79.
22. Mobasheri A, Shakibaei M, Marples D. Immunohistochemical localization of aquaporin 10 in the apical membranes of the human ileum: a potential pathway for luminal water and small solute absorption. Histochem Cell Biol 2004;121:463-71.
23. Kaminsky LS, Zhang QY. The small intestine as a xenobiotic-metabolizing organ. Drug Metab Dispos 2003;31:1520-5.
24. Langmann T, Schmitz G. Loss of detoxification in inflammatory bowel disease. Nat Clin Pract Gastroenterol Hepatol 2006;3:358-9.
25. Catassi C, Bearzi I, Holmes GK. Association of coeliac disease and intestinal lymphomas and other cancers. Gastroenterology 2005;128:S79-86.
26. Wice BM, Gordon JI. A tetraspan membrane glycoprotein produced in the human intestinal epithelium and liver that can regulate cell density-dependent proliferation. J Biol Chem 1995;270:21907-18.
27. Diosdado B, Wapenaar MC, Franke L, et al. A microarray screen for novel candidate genes in coeliac disease pathogenesis. Gut 2004;53:944-51.
28. Wahab PJ, Crusius JB, Meijer JW, et al. Gluten challenge in borderline gluten-sensitive enteropathy. The American journal of gastroenterology 2001;96:1464-9.
29. Dring MM, Goulding CA, Trimble VI, et al. The pregnane X receptor locus is associated with susceptibility to inflammatory bowel disease. Gastroenterology 2006;130:341-8; quiz 592.
30. MacDonald TT. Epithelial proliferation in response to gastrointestinal inflammation. Ann N Y Acad Sci 1992;664:202-9.

Supplementary table 1. Details on coeliac disease patients and controls included in this study

Individual ID	Diagnosis	Marsh Stage	GFD	Age (yrs)	Gender	HLA
223	CD	MIII	no	67	F	DQ2/8
193	CD	MIII	no	39	M	DQ2/X
259	CD	MIII	no	54	F	DQ2/2
100	CD	MIII	no	16	F	DQ2/2
180	CD	MIII	no	2	M	DQ2/X
310	CD	MIII	no	78	F	DQ2/X
276	CD	MIII	no	33	F	DQ2/X
284	CD	MIII	no	1	F	DQ2/X
221	CD	MIII	no	37	F	DQ2/X
341	CD	MIII	no	19	F	DQ2/2
375	CD	MIII	no	62	M	DQ2/X
369	CD	MIII	no	17	F	DQ2/X
330	CD	MIII	no	2	M	DQ2/2
286	CD	MIII	yes	70	M	DQ2/X
406	CD	MIII	yes	55	F	DQ2/2
230	CD	MIII	yes	60	F	DQ2/X
109	CD	MII	yes	35	F	ND
389	CD	MII	yes	30	F	DQ2/X
54	CD	MII	yes	43	F	DQ2/X
248	CD	MII	yes	36	F	DQ2/X
260	CD	MII	yes	59	M	ND
194	CD	MII	yes	52	M	DQX/X
408	CD	MII	yes	49	F	DQ2/8
185	CD	MII	yes	31	F	DQ2/X
217	CD	MII	yes	51	F	ND
343	CD	MI	yes	59	M	ND
88	CD	MI	yes	50	M	DQ2/8
231	CD	MI	yes	19	F	DQ2/2
202	CD	M0-I	yes	64	M	DQ2/8
237	CD	M0-I	yes	55	F	DQ2/X
201	CD	M0-I	yes	35	F	DQ2/2
110	CD	M0-I	yes	52	F	DQ2/2
192	CD	M0-I	yes	50	M	DQ2/X
368	CD	M0	yes	2	F	DQ2/X
225	CD	M0	yes	89	F	DQ2/X
190	CD	M0	yes	28	F	DQ2/8
321	CD	M0	yes	82	F	DQ2/X
339	CD	M0	yes	60	F	DQ2/X
105	CD	M0	yes	75	F	DQ2/2

Individual ID	Diagnosis	Marsh Stage	GFD	Age (yrs)	Gender	HLA
218	CD	M0	yes	70	F	ND
216	CD	M0	yes	40	F	DQ2/X
246	CD	M0	yes	27	M	DQ2/2
308	CD	M0	yes	53	F	DQ2/X
178	CD	M0	yes	25	M	DQ2/2
208	Control	N/A	N/A	70	F	ND
188	Control	N/A	N/A	39	F	ND
285	Control	N/A	N/A	37	F	ND
385	Control	N/A	N/A	54	F	ND
332	Control	N/A	N/A	54	F	ND
158	Control	N/A	N/A	31	F	DQX/X
102	Control	N/A	N/A	30	F	DQ2/8
388	Control	N/A	N/A	4	F	DQX/X
200	Control	N/A	N/A	35	F	DQ2/X
220	Control	N/A	N/A	15	F	ND
376	Control	N/A	N/A	22	F	ND
Ams9	Control	N/A	N/A	35	F	ND
331	Control	N/A	N/A	19	F	ND
Ams 5	Control	N/A	N/A	32	F	ND
247	Control	N/A	N/A	45	M	ND
104	Control	N/A	N/A	45	F	DQX/X
338	Control	N/A	N/A	36	F	ND
256	Control	N/A	N/A	50	M	ND
210	Control	N/A	N/A	27	F	DQX/X
239	Control	N/A	N/A	59	F	ND
196	Control	N/A	N/A	20	F	ND

Abbreviations (in order of appearance): GFD, gluten-free diet; CD, coeliac disease; ND: not determined; N/A, not applicable.

SECTION III

MOLECULAR PATHWAY ANALYSIS

Chapter 6

A transcriptome of the tight junction gene network in the normal and coeliac disease-affected duodenal mucosa

MC Wapenaar, M van Kollenburg, CJJ Mulder, and C Wijmenga

(Submitted for publication)

ABSTRACT

Background & Aims: Coeliac disease (CD), a gluten-induced malabsorption disorder, is characterized by enhanced intestinal permeability due to a dysfunctional tight junction (TJ). Our aim was to make a comprehensive transcriptional analysis of genes involved in the TJ network, both from the normal and the CD-affected duodenal mucosa. ***Methods:*** In total, 93 TJ-related genes were examined for transcriptional activity by qRT-PCR on low-density arrays. RNA was extracted from duodenal biopsies derived from non-CD controls (n=14), CD patients in complete remission (n=16), and patients with atrophic CD (n=18). ***Results:*** Transcriptional profiles were made from TJ genes coding for transmembrane proteins, adaptors, signal transducers, transcriptional regulators, and cytoskeletal proteins. The claudin profile was dominated by *CLDN15, -3*, and *–7* and was evolutionary conserved in rodents. Three pairs of claudins revealed tight transcriptional co-regulation. In atrophic CD patients, the claudin profiles revealed a dynamic picture marked by extreme outliers, particularly for *CLDN10*. In 15 out of 18 cases, patients showed at least one claudin with an outlier-value, possibly causing a distorted claudin composition. The actin-binding *ACTN4*, and actin-regulator *RHOA*, were highly expressed and showed increased expression. We also observed remarkable induction of the growth factor *EGF* (> 60-fold in atrophic CD), which has a preserving effect on the epithelial barrier. ***Conclusions:*** The TJ transcriptome of the duodenum that we presented here lays a firm foundation for further genomics and proteomics research in intestinal disorders. The TJ expression profile in atrophic CD points toward a distorted claudin regulation, and changed dynamics of the actinomyosin cytoskeleton.

INTRODUCTION

Coeliac disease (CD) is a gluten-induced inflammatory condition of the small intestine characterized by atrophy of the villi, crypt hyperplasia, and lymphocyte infiltration[1]. The associated nutrient absorption deficiency results in a complex of clinical features[2]. The inflammation also has a negative effect on the barrier function of the epithelium as is evident from the increased permeability for macromolecules[3], the reduced trans-epithelial resistance[4], and the induced immune response to luminal commensals like *Saccharomyces cerevisiae*[5, 6].

The tight junction (TJ) is the most apical epithelial junction and plays a key role in selective paracellular transport, barrier function, and polarity[7]. A functional TJ consists of transmembrane proteins and membrane-associated adaptors that relay information to an intricate network of signal transducers and transcriptional regulators[8-10]. Although enhanced permeability in CD is an established fact for over 30 years, little is known about the mechanism behind this phenomenon. Indeed, most permeability studies in CD are limited to observations on the TJ

adaptor protein ZO-1[11]. In general, information is scarce on the transcriptional regulation of TJ-related genes in the human small intestine. Here we present a comprehensive gene expression study of the extended TJ gene network in both the healthy duodenum, and in the duodenal mucosa affected by CD. For the first time this transcriptome gives insight in the TJ gene activity in the normal intestine and the complex transcriptional dynamics resulting from the gluten-induced inflammation.

MATERIALS AND METHODS

Patient material

This study was performed with 14 non-CD controls, 16 CD patients in complete remission (Marsh 0, M0) on a gluten-free diet, and 18 atrophic CD patients (Marsh III, MIII). Duodenal biopsy samples were collected with forced forceps endoscopy. Histological examination and Marsh classification, according to the UEWG criteria[12], were performed by experienced pathologists (GAM and JWRM). Biopsies to be used for RNA isolation were stored in either liquid nitrogen, or RNALater (Ambion, Austin, TX) at –20°C. Patients enrolled in the study signed an informed consent. This study was approved by the Medical Ethics Committee of the University Medical Center Utrecht.

Gene expression

The isolation of total RNA from duodenal biopsies, using tissue homogenization with glass beads and extraction with TRIzol (Invitrogen, Carlsbad, CA), was as described previously[13] (Wapenaar, 2004). All subsequent steps were performed with reagents, equipment, and protocols purchased from Applied Biosytems (Foster City, CA). RNA samples of 1 ug were used for cDNA synthesis with the High Capacity cDNA Archive Kit. Pools were made from individually prepared cDNA samples for the 14 non-CD controls and the 16 CD M0 patients. The 18 CD MIII samples were tested separately. Custom-made Low Density Arrays, LDA (formerly known as Micro Fluidic Cards), were loaded with cDNA from 800 ng RNA, suspended in 800 ul Universal Master Mix, and equally separated over the eight channels of each card. The custom LDAs consisted of 384-well cards loaded with Taqman assays for 93 TJ target genes and the endogenous control gene *GUSB*, all spotted in quadruplicate. Target genes and assay catalogue numbers are summarized in table 1. PCR cycling and data collection was performed on an SDS 7900HT. SDS2.2 software was used for data analysis. Quantification of gene expression in the normal intestine was relative to that of the *GUSB* gene. Gene expression in CD M0 and MIII samples was normalized to that of the non-CD controls.

RESULTS

The TJ gene network

We based our gene list of the TJ pathway on reports in the literature[8-10], and the KEGG database (www.genome.jp/kegg/). The genes in this network, and the known interactions of their products, are depicted in figure 1.

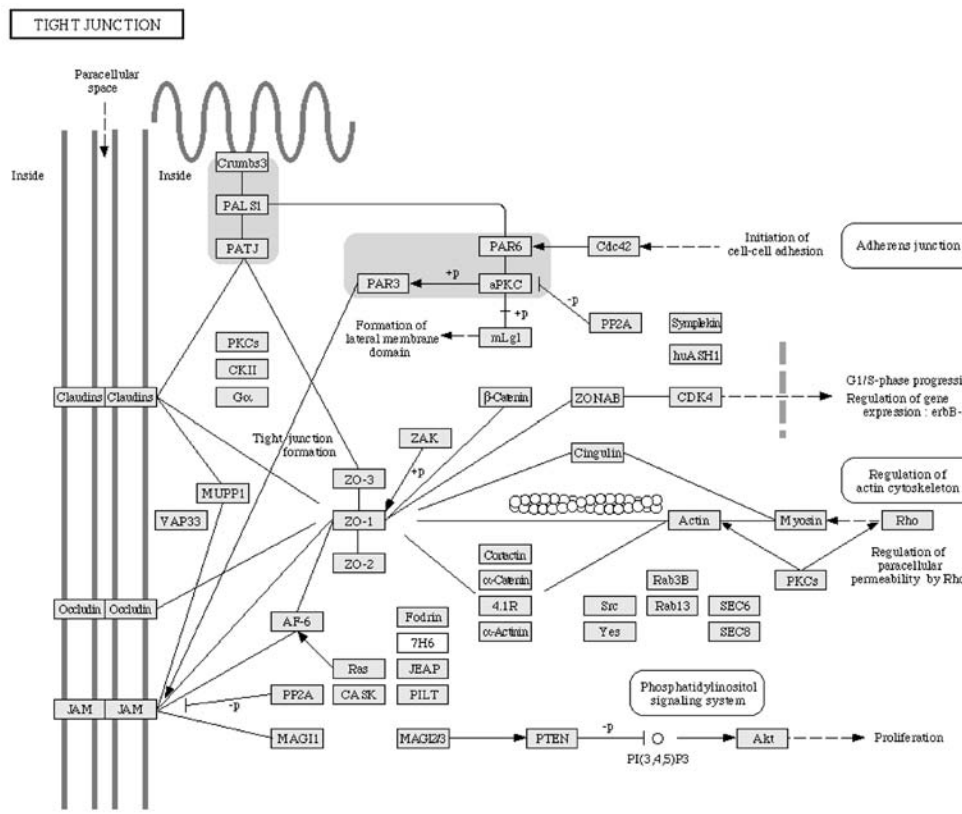


Figure 1. Interactive gene network related to TJ functionality, according to KEGG.

The normal duodenum

The expression of 93 TJ genes in the normal duodenal mucosa relative to the endogenous control *GUSB* is depicted in figure 2, and summarized in table 1. The expression profile of the 25 tested transmembrane genes is dominated by *CLDN15* and, to a lesser extent, by *CLDN3* and *CLDN7* (figure 2A). Ordered according to the expression level: *CLDN15* > *-3* > *-7* > *-12* > *-4* > *-2* > *-17* > *-8* > *-10* > *-5* > *-6* > *-9* > *-18* > *-1* > *-11*. *CLDN14*, *-16*, *-19*, and *–20* were below the detection level. Of the *crumbs* family, only *CRB3* appeared to be expressed in the duodenum, at a similar level as *occludin*. The *JAM-1* gene *F11R* was moderately high expressed, in contrast

to the *JAM-4* gene *LOC150084*, that was hardly detectable (figure 2A). The adaptor gene profile (n=21) is shown in figure 2B and is characterized by a high expression for the occluding-binding and vesicle targeting *VAPA* gene[14]. Moderately high

Table 1. Expression of TJ pathway genes in the normal duodenal mucosa, relative to the endogenous control gene *GUSB*

Symbol	Alias	Assay ID	RQ	minus SD	plus SD
Transmembrane proteins					
CLDN1		Hs00221623_m1	0.0179	0.0160	0.0202
CLDN2		Hs00252666_s1	0.5542	0.5303	0.5792
CLDN3		Hs00265816_s1	4.5996	4.4181	4.7886
CLDN4		Hs00533616_s1	0.7531	0.7125	0.7959
CLDN5		Hs00533949_s1	0.1389	0.1301	0.1482
CLDN6		Hs00607528_s1	0.0682	0.0601	0.0774
CLDN7		Hs00600772_m1	2.1317	2.0575	2.2085
CLDN8		Hs00273282_s1	0.2752	0.2596	0.2918
CLDN9		Hs00253134_s1	0.0478	0.0435	0.0525
CLDN10		Hs00199599_m1	0.1909	0.1467	0.2484
CLDN11		Hs00194440_m1	0.0134	0.0109	0.0164
CLDN12		Hs00273258_s1	0.7607	0.7107	0.8142
CLDN14		Hs00377953_m1	<0.0003		
CLDN15		Hs00204982_m1	7.9086	7.6137	8.2148
CLDN16		Hs00198134_m1	<0.0003		
CLDN17		Hs00273276_s1	0.3607	0.3434	0.3789
CLDN18		Hs00212584_m1	0.0293	0.0274	0.0314
CLDN19		Hs00381204_m1	<0.0003		
CLDN20		Hs00378662_m1	<0.0003		
CRB1		Hs00201372_m1	<0.0003		
CRB2		Hs00543624_m1	<0.0003		
CRB3		Hs00373616_m1	0.5854	0.5633	0.6085
F11R	*JAM-1*	Hs00375889_m1	2.3432	2.2234	2.4695
LOC150084	*JAM-4*	Hs00401212_m1	0.0103	0.0094	0.0112
OCLN		Hs00170162_m1	0.4715	0.4487	0.4955
Adaptors					
CASK		Hs00177620_m1	2.6723	2.4301	2.9387
CGN		Hs00430426_m1	2.6989	2.5141	2.8973
CXADR	*CAR*	Hs00154661_m1	2.1378	2.0393	2.2411
INADL	*PATJ*	Hs00195106_m1	0.9301	0.8787	0.9845
LLGL1		Hs00188098_m1	0.0657	0.0582	0.0743
LLGL2		Hs00189729_m1	2.3512	2.2379	2.4703
MAGI1	*AIP3*	Hs00191026_m1	0.5800	0.5421	0.6205
MAGI2	*AIP1*	Hs00202321_m1	0.0459	0.0313	0.0674

Symbol	Alias	Assay ID	RQ	minus SD	plus SD
MAGI3		Hs00326365_m1	0.3029	0.2922	0.3139
MPDZ	*MUPP1*	Hs00187106_m1	0.0700	0.0603	0.0812
MPP5	*PALS1*	Hs00223885_m1	1.2986	1.1863	1.4215
PARD3	*PAR3*	Hs00219744_m1	0.2537	0.2377	0.2708
ALS2CR19	*PAR3B*	Hs00365140_m1	0.1907	0.1729	0.2103
PARD6A	*PAR6*	Hs00180947_m1	0.0213	0.0195	0.0234
PARD6B	*PAR6B*	Hs00325996_m1	0.2443	0.2276	0.2622
PARD6G	*PAR-6G*	Hs00261284_m1	0.0138	0.0103	0.0186
TJP1	*ZO-1*	Hs00268480_m1	0.8703	0.7981	0.9491
TJP2	*ZO-2*	Hs00178081_m1	0.0155	0.0122	0.0198
TJP3	*ZO-3*	Hs00274276_m1	0.9380	0.8551	1.0289
TJAP1	*PILT*	Hs00369777_m1	0.0371	0.0315	0.0438
VAPA	*VAP33*	Hs00427749_m1	5.5772	5.1873	5.9964
Signalling proteins					
AMOTL1	*JEAP*	Hs00294070_m1	0.2134	0.1954	0.2331
ARHGEF2	*GEF-H1*	Hs00190884_m1	0.1381	0.1092	0.1746
CDC42		Hs00741586_mH	0.0101	0.0076	0.0134
CSNK2A1		Hs00601957_m1	0.1152	0.1046	0.1268
MLLT4	*AF6*	Hs00180592_m1	1.0994	1.0035	1.2044
PPM1J		Hs00374513_m1	0.0069	0.0047	0.0101
PPP2CA		Hs00427259_m1	0.8939	0.8302	0.9625
PPP2CB		Hs00602137_m1	1.2641	1.1769	1.3577
PPP2R1A		Hs00204426_m1	2.8818	2.7842	2.9828
PPP2R1B		Hs00184737_m1	0.5352	0.4687	0.6111
PPP2R2A		Hs00160392_m1	1.0793	1.0252	1.1362
PPP2R2B		Hs00270227_m1	0.0737	0.0658	0.0826
PPP2R2C		Hs00739033_m1	0.0280	0.0256	0.0305
PPP2R3A		Hs00160407_m1	0.0993	0.0939	0.1050
PPP2R3B		Hs00203045_m1	0.1005	0.0912	0.1108
PPP2R4		Hs00603515_m1	0.4484	0.4327	0.4646
PRKCA		Hs00176973_m1	0.5701	0.5511	0.5897
PRKCB1		Hs00176998_m1	0.0916	0.0841	0.0997
PRKCD		Hs00178914_m1	1.2632	1.2100	1.3188
PRKCE		Hs00178455_m1	0.1890	0.1740	0.2054
PRKCG		Hs00177010_m1	0.0263	0.0236	0.0293
PRKCH		Hs00178933_m1	0.4629	0.4493	0.4769
PRKCI		Hs00702254_s1	0.0419	0.0374	0.0469
PRKCQ		Hs00234697_m1	0.0907	0.0790	0.1040
PRKCZ		Hs00177051_m1	1.3055	1.2291	1.3866
PTEN		Hs00829813_s1	0.0607	0.0526	0.0700
RAB3B		Hs00267896_m1	0.0498	0.0426	0.0583
RAB13		Hs00762784_s1	0.3537	0.3085	0.4054

Symbol	Alias	Assay ID	RQ	minus SD	plus SD
RAC1		Hs00251654_m1	0.1810	0.1639	0.1999
RHOA		Hs00357608_m1	7.9990	7.5897	8.4304
ROCK1		Hs00178463_m1	0.4844	0.4689	0.5004
ROCK2		Hs00178154_m1	1.2579	1.1926	1.3268
SEC6L1		Hs00292654_m1	0.3544	0.3424	0.3667
SEC8L1		Hs00253986_m1	1.2648	1.1829	1.3523
ZAK	*MLK7*	Hs00370447_m1	0.1009	0.0915	0.1113
Actin-binding					
ACTN1		Hs00241650_m1	0.5701	0.5531	0.5876
ACTN4		Hs00245168_m1	5.1159	4.8622	5.3828
CTTN		Hs00193322_m1	1.0674	0.9587	1.1886
EPB41		Hs00270045_m1	0.9752	0.9047	1.0513
MLCK		Hs00417542_m1	<0.0003		
MYLK		Hs00364926_m1	0.4206	0.3836	0.4611
SPTAN1	*alpha fodrin*	Hs00162203_m1	2.2274	1.9715	2.5165
Transcriptional regulators					
ASH1L	*huASH1*	Hs00218516_m1	1.0190	0.9730	1.0672
CDK4		Hs00364847_m1	0.8909	0.8345	0.9511
SYMPK		Hs00191361_m1	0.5579	0.5268	0.5909
Growth factor					
EGF		Hs00153181_m1	<0.0003		
EGFR		Hs00193306_m1	0.1658	0.1510	0.1819

Abbreviations: RQ, relative quantification; SD, standard deviation from quadruplicate measurements.

expression was found for *CASK*, *CGN*, *LLGL2*, and *CXADR (CAR)*, and moderate expression for *MPP5 (PALS1)*, *INADL (PATJ)*, *TJP1 (ZO-1)*, and *TJP3 (ZO-3)*. *TJP2* (ZO-2) was very low expressed. The MAGUK family members *MAGI1* > *MAGI3* > *MAGI2* had expression levels in the order as indicated. The signal transduction profile (n=35) was marked by a high expression of the small GTPase *RHOA*, and a moderately high expression for the protein phosphatase 2, regulatory subunit *PPP2R1A* (figure 2C). Moderate expression was observed for other *PPP2*-related genes, *PRKCD*, *PRKCZ*, *MLLT4* (*AF6*), *ROCK2*, and *SEC8L1*. Figure 2D shows the expression of seven actin-binding genes in which *ACTN4* stands out. Three transcriptional regulators (*ASH1L*, *CDK4*, and *SYMPK*) are moderate expressed. The growth factor *EGF*, known for its barrier-promoting effect, was below detection level. Its receptor *EGFR* showed low-level expression.

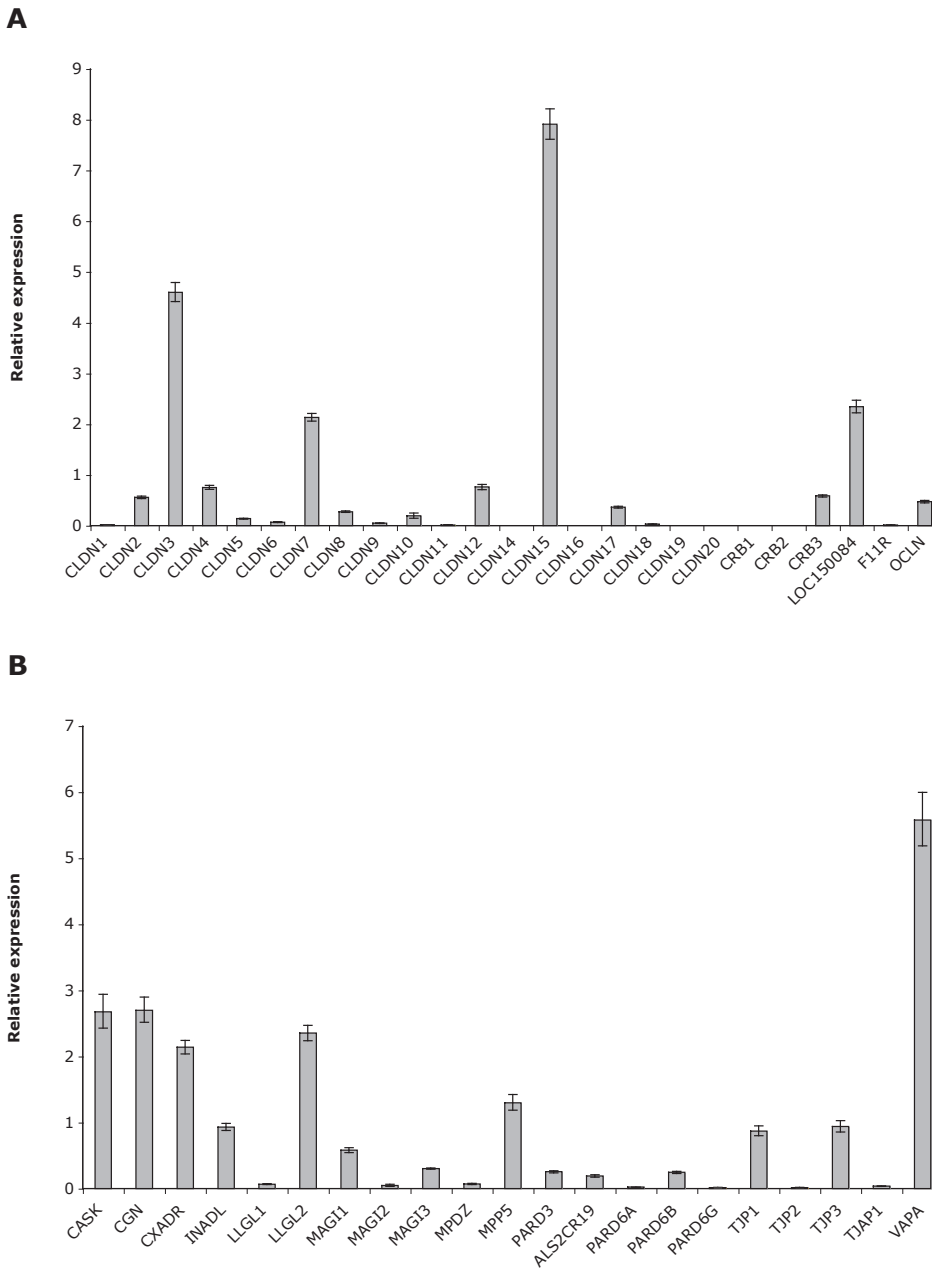
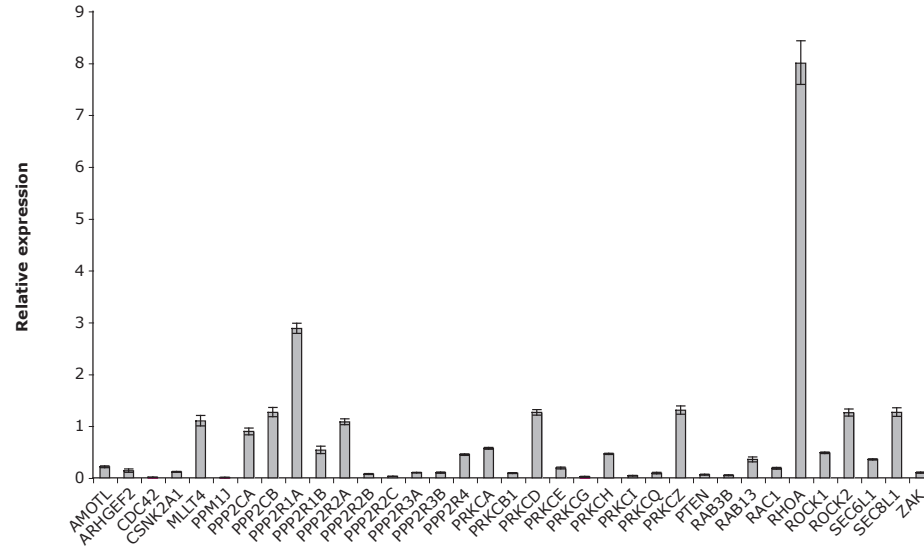
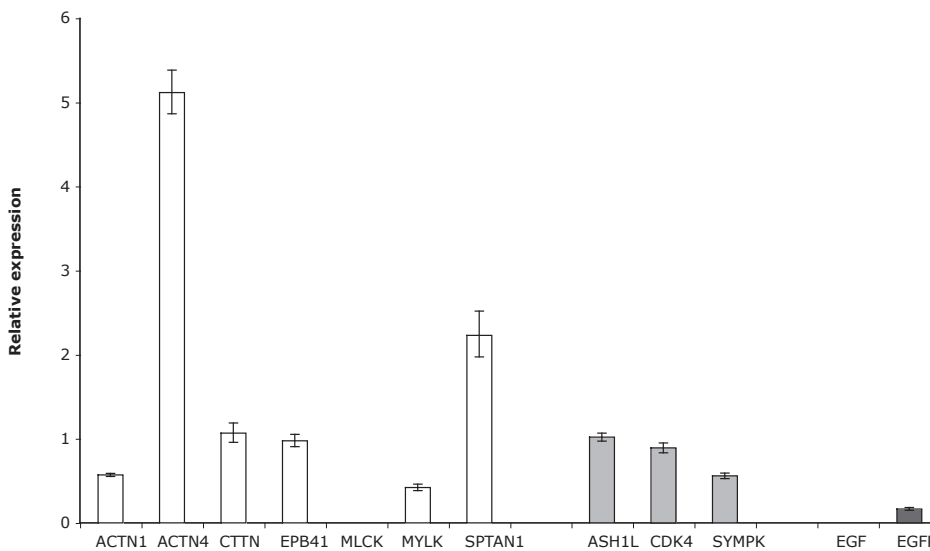


Figure 2. Quantification of the transcriptional activity of TJ pathway genes in the normal duodenal mucosa, relative to the housekeeping gene *GUSB*. Genes are grouped according to the function of their products: **A**) structural transmembrane components of the TJ; **B**) adaptors; **C**) signal transduction; **D**) actin-binding proteins (blanc), transcriptional regulators (grey), and growth factors (black). Error bars indicate standard deviations from the average of quadruplicate measurements.

C


Relative expression
0
1
2
3
4
5
6
7
8
9
AMOTL
ARHGEF2
CDC42
CSNK2A1
MLLT4
PPM1J
PPP2CA
PPP2CB
PPP2R1A
PPP2R1B
PPP2R2A
PPP2R2B
PPP2R2C
PPP2R3A
PPP2R3B
PPP2R4
PRKCA
PRKCB1
PRKCD
PRKCE
PRKCG
PRKCH
PRKCI
PRKCQ
PRKCZ
PTEN
RAB3B
RAB13
RAC1
RHOA
ROCK1
ROCK2
SEC6L1
SEC8L1
ZAK

D


Relative expression
0
1
2
3
4
5
6
ACTN1
ACTN4
CTTN
EPB41
MLCK
MYLK
SPTAN1
ASH1L
CDK4
SYMPK
EGF
EGFR

The coeliac disease-affected duodenum

Next, we assessed the profile of the same TJ genes expressed in the duodenal mucosa of CD patients. We determined the expression in pooled cDNA from 16 patients in complete remission (M0), as well as 18 individual CD cases with villous atrophy (MIII), relative to that in pooled cDNA from 14 non-CD controls. These results are summarized in table 2. The members of the claudine family displayed a remarkable dynamic expression behavior leading to considerable expression differences among the MIII patients. In the most extreme case this could amount to a 3000-fold difference (*CLDN10*, patient 2 versus 6). The claudins most variable in expression were *CLDN6*, *-8*, *-10*, *-17*, and *–18* (CV > 100). This resulted in outliers with an extreme low or high expression. *CLDN10* expression was particularly variable as seven out of 18 MIII cases were low outliers (< 0.05-fold), with a single high outlier (3.32-fold). *CLDN8*, *-17*, and-*18*, showed five, four, and four low outliers, respectively. We also observed that in some of the patients more outlier data accumulated, e.g. patient 3 showed four claudins with extreme low (*CLDN8*, *-10*, *17*, and *-18*), and a single claudin (*CLDN5*) with high expression. In contrast, patient 16 and 18 showed high outliers for five and seven claudins, respectively. In total, twelve out of the 18 MIII cases showed for at least one claudin extreme low expression. Considering all MIII cases, there were only three patients (cases 4, 7, and 9) in which not a single claudin displayed extreme expression behavior. Because of these strong expression fluctuations among the patients it was not possible to make a general remark about up- or down-regulation of specific claudins as a result of the CD pathology.

Transcriptional co-regulation of claudin genes

When we plotted the relative expression of the claudin genes for all 18 MIII patients we observed that some patterns appeared similar, suggesting co-regulation of gene expression (figure 3). We made a pair-wise comparison for all the expression data of the 15 claudins expressed in the duodenum and calculated the correlation coefficients. There were five pairs of claudins with a high correlation ($r^2 > 0.7$) in expression (figure 3A). These could be grouped in two basic profiles (figure 3B and 3C) of three highly correlated claudin pairs (*CLDN2* and *–12*; *CLDN6* and *–9*; and *CLDN8* and *-17*). CLDN6 and –9 are juxtapositioned in a tail-to-tail fashion on chromosome 16p13.3, while CLDN8 and –17 are adjacent tail-to-head on chromosome 21q22.11. A pair-wise location in the genome does not necessarily lead to co-regulation of claudins as exemplified by *CLD3* and *CLDN4*, which are organized head-to-head on chromosome 7q11.23 but lack expression correlation (figure 3A). In contrast, *CLDN2* and *CLDN12* are highly co-regulated

Table 2. Expression of TJ pathway genes in the duodenal mucosa of CD-affected atrophic stage

Symbol	Alias	Assay ID	RQ CD M0	1	2	3	4	5	6	7
Transmembrane proteins										
CLDN1		Hs00221623_m1	1.63	2.00	1.85	1.76	1.21	5.69	3.36	0.50
CLDN2		Hs00252666_s1	0.96	1.47	2.89	1.07	2.23	2.16	1.73	1.08
CLDN3		Hs00265816_s1	1.33	0.85	1.16	0.97	1.44	1.12	0.87	0.90
CLDN4		Hs00533616_s1	0.92	1.36	1.05	0.88	0.94	1.25	0.86	0.70
CLDN5		Hs00533949_s1	1.01	0.43	0.56	4.54	0.62	0.74	0.53	1.96
CLDN6		Hs00607528_s1	1.26	0.35	0.07	0.64	0.59	0.89	0.43	0.58
CLDN7		Hs00600772_m1	1.02	0.92	1.63	1.77	1.15	1.48	0.97	0.88
CLDN8		Hs00273282_s1	1.71	0.09	0.03	0.08	0.31	0.26	0.24	0.24
CLDN9		Hs00253134_s1	1.22	0.24	0.37	0.84	0.53	0.74	0.61	0.96
CLDN10		Hs00199599_m1	0.89	0.98	3.32	0.00	2.47	0.15	0.00	0.43
CLDN11		Hs00194440_m1	2.06	0.74	1.51	1.95	1.25	3.16	1.91	1.51
CLDN12		Hs00273258_s1	1.30	0.95	1.42	0.63	0.96	1.54	1.25	1.27
CLDN14		Hs00377953_m1	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.
CLDN15		Hs00204982_m1	1.13	0.49	1.08	1.27	1.05	0.95	0.78	0.86
CLDN16		Hs00198134_m1	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.
CLDN17		Hs00273276_s1	2.10	0.09	0.03	0.06	0.28	0.35	0.33	0.28
CLDN18		Hs00212584_m1	1.87	2.79	3.21	0.53	4.27	1.09	0.98	1.05
CLDN19		Hs00381204_m1	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.
CLDN20		Hs00378662_m1	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.
CRB1		Hs00201372_m1	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.
CRB2		Hs00543624_m1	>89.7	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	>27.4
CRB3		Hs00373616_m1	1.00	0.58	1.32	1.02	0.63	1.38	0.81	0.52
F11R	*JAM-1*	Hs00375889_m1	1.07	0.61	0.98	1.12	0.92	1.37	0.80	0.82
LOC150084	*JAM-4*	Hs00401212_m1	1.34	0.39	0.79	1.70	0.77	1.12	0.56	0.49
OCLN		Hs00170162_m1	1.31	0.74	1.49	0.98	1.10	1.30	0.94	1.03
Adaptors										
CASK		Hs00177620_m1	1.31	0.74	1.49	0.98	1.10	1.30	0.94	1.03
CGN		Hs00430426_m1	0.95	0.55	0.78	0.82	0.67	0.94	0.51	0.59
CXADR	*CAR*	Hs00154661_m1	1.37	0.54	1.15	0.90	0.93	1.40	0.87	1.12
INADL	PATJ	Hs00195106_m1	1.32	0.77	1.25	1.12	1.32	1.45	1.23	1.19
LLGL1		Hs00188098_m1	1.02	0.72	0.99	1.92	1.02	1.25	0.92	1.12
LLGL2		Hs00189729_m1	1.03	0.82	1.16	1.26	1.19	1.07	0.95	0.76
MAGI1	AIP3	Hs00191026_m1	1.06	0.44	0.84	0.93	0.73	0.74	0.54	0.69
MAGI2	*AIP1*	Hs00202321_m1	1.02	0.00	0.57	0.58	1.39	0.84	0.65	0.67
MAGI3		Hs00326365_m1	1.11	0.55	1.50	1.34	0.73	1.31	0.97	1.03
MPDZ	*MUPP1*	Hs00187106_m1	1.12	0.43	0.87	1.11	0.67	1.36	0.83	1.48

MIII, and remission stage M0, relative to non-CD controls

RQ CD MIII patients											Avrg	SD	CV
8	9	10	11	12	13	14	15	16	17	18			
5.42	1.15	0.66	1.70	1.70	2.75	1.07	1.00	4.49	1.05	5.88	2.40	1.79	75
2.10	0.98	0.98	1.84	2.49	1.05	1.58	0.78	3.54	1.29	5.57	1.93	1.17	61
1.28	1.98	2.54	0.91	0.82	1.36	1.82	1.85	1.27	1.79	0.97	1.33	0.49	37
0.32	0.79	0.69	0.72	0.83	0.55	0.85	0.70	1.03	1.22	1.83	0.92	0.34	37
0.25	0.97	0.49	0.49	0.83	1.51	1.38	0.86	1.05	1.05	0.82	1.06	0.97	91
0.59	1.65	1.14	0.35	0.46	1.87	3.08	1.34	3.37	1.50	6.88	1.43	1.64	114
0.53	0.86	0.68	1.21	1.14	0.94	0.99	0.47	1.25	1.09	1.35	1.07	0.35	32
1.07	0.74	0.94	0.05	0.15	0.97	2.50	0.61	7.29	0.81	7.19	1.31	2.24	171
0.34	1.80	1.67	0.45	0.55	1.84	1.90	1.32	2.41	1.04	3.26	1.16	0.83	72
1.42	0.15	0.02	0.26	0.31	0.02	0.04	0.02	1.98	0.01	1.06	0.70	0.99	141
0.96	1.26	0.97	0.59	1.81	3.86	1.61	1.39	2.61	1.06	1.94	1.67	0.84	50
1.24	1.09	0.75	1.26	1.38	1.16	1.30	0.57	3.19	1.21	3.63	1.38	0.79	57
b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.			
0.28	1.14	0.72	0.76	0.89	0.78	0.61	0.71	0.96	1.18	0.77	0.85	0.25	29
>8.0	b.d.	b.d.	>9.6	b.d.	b.d.	b.d.	b.d.	>18.2	b.d.	>18.3			
0.95	1.07	0.99	0.06	0.24	1.05	3.08	0.64	10.17	0.87	9.62	1.68	3.07	183
36.82	5.27	0.29	5.12	1.20	0.50	0.75	0.78	16.22	0.45	22.61	5.77	9.78	169
b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.			
b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.			
b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	>9.41	b.d.	b.d.			
>61.7	>9.9	b.d.	>219	b.d.	>684	b.d.	b.d.	b.d.	>447	b.d.			
0.40	0.94	0.73	1.14	0.88	0.49	0.33	0.50	1.06	0.81	0.79	0.80	0.30	38
0.69	1.03	0.72	0.97	0.96	0.96	0.98	1.05	1.34	1.08	0.88	0.96	0.20	21
0.01	0.54	0.07	0.58	0.05	1.26	1.77	0.11	0.94	1.12	0.68	0.72	0.53	73
1.32	1.24	0.77	0.77	0.71	0.84	0.92	0.65	1.72	1.38	0.85	1.04	0.30	29
1.32	1.24	0.77	0.77	0.71	0.84	0.92	0.65	1.72	1.38	0.85	1.04	0.30	29
0.33	0.74	0.74	0.74	0.70	0.80	1.08	0.63	0.81	1.40	0.89	0.76	0.23	31
0.83	1.95	0.72	0.81	0.98	0.72	0.77	0.87	1.82	1.26	0.55	1.01	0.39	39
1.36	1.44	0.94	1.03	1.05	1.38	1.29	1.11	1.70	1.46	1.16	1.24	0.22	18
0.71	0.79	0.60	1.10	1.35	1.77	1.02	0.92	1.12	1.12	0.88	1.07	0.34	32
0.57	0.95	1.24	0.90	0.91	1.38	1.93	1.30	0.98	1.43	1.37	1.12	0.31	28
0.47	0.88	0.76	0.57	0.53	0.94	0.90	1.09	0.84	1.05	0.59	0.75	0.20	26
0.46	0.88	0.32	0.79	1.26	1.47	1.91	2.12	1.57	0.62	1.23	0.96	0.57	59
0.82	1.18	0.56	0.86	0.91	0.71	0.55	0.60	1.29	0.91	0.74	0.92	0.30	32
0.68	1.11	0.49	0.68	1.25	1.48	1.00	0.71	1.71	0.99	0.58	0.97	0.37	39

Symbol	Alias	Assay ID	RQ CD M0							
				1	2	3	4	5	6	7
MPP5	PALS1	Hs00223885_m1	1.30	0.50	0.93	0.85	1.06	0.81	0.72	0.83
PARD3	PAR3	Hs00219744_m1	0.97	0.71	0.97	1.01	1.09	1.15	0.97	1.14
ALS2CR19	PAR3B	Hs00365140_m1	1.33	0.57	1.09	1.79	0.99	1.21	1.01	1.22
PARD6A	PAR6	Hs00180947_m1	0.97	1.05	1.04	1.85	0.51	1.55	1.36	0.80
PARD6B	PAR6B	Hs00325996_m1	1.19	0.47	0.64	0.76	0.75	0.78	0.75	0.75
PARD6G	PAR-6G	Hs00261284_m1	1.94	0.86	1.24	3.01	1.37	0.77	1.24	3.19
TJP1	ZO-1	Hs00268480_m1	1.10	0.46	0.97	0.66	1.03	0.95	0.68	0.79
TJP2	ZO-2	Hs00178081_m1	0.63	0.51	0.50	2.53	0.82	0.83	1.03	1.60
TJP3	ZO-3	Hs00274276_m1	0.91	0.51	1.13	1.39	1.00	0.80	0.65	0.59
TJAP1	PILT	Hs00369777_m1	1.19	0.69	1.10	1.28	0.94	1.65	1.24	1.17
VAPA	VAP33	Hs00427749_m1	1.09	0.59	0.98	0.83	0.91	1.15	0.93	0.82
Signalling proteins										
AMOTL1	JEAP	Hs00294070_m1	1.02	0.36	0.69	0.80	1.16	1.01	0.66	0.91
ARHGEF2	GEF-H1	Hs00190884_m1	1.03	0.66	0.69	0.83	1.14	1.24	1.12	1.04
CDC42		Hs00741586_mH	1.63	1.70	0.87	0.85	2.26	1.29	1.65	1.02
CSNK2A1		Hs00601957_m1	1.09	1.11	1.08	1.26	1.66	1.40	0.97	1.13
MLLT4	AF6	Hs00180592_m1	1.13	0.61	1.03	0.98	0.95	1.18	0.72	0.85
PPM1J		Hs00374513_m1	1.32	1.56	2.19	4.79	1.71	0.98	1.93	2.13
PPP2CA		Hs00427259_m1	1.39	1.11	1.63	1.39	1.47	2.02	1.62	1.28
PPP2CB		Hs00602137_m1	1.13	0.92	1.49	1.46	1.05	2.15	1.43	0.98
PPP2R1A		Hs00204426_m1	0.93	0.79	1.45	2.09	1.16	1.30	1.01	0.92
PPP2R1B		Hs00184737_m1	1.17	0.96	1.60	1.67	1.11	1.79	1.60	1.10
PPP2R2A		Hs00160392_m1	1.24	0.89	1.40	1.41	1.40	1.81	1.70	1.49
PPP2R2B		Hs00270227_m1	1.14	0.77	0.72	0.89	1.07	0.84	1.84	1.16
PPP2R2C		Hs00739033_m1	0.57	0.62	0.88	0.03	0.59	0.28	0.21	1.49
PPP2R3A		Hs00160407_m1	0.98	0.32	0.77	0.62	0.73	0.62	0.33	0.59
PPP2R3B		Hs00203045_m1	0.94	0.65	1.24	1.62	1.02	0.94	1.16	0.68
PPP2R4		Hs00603515_m1	1.09	0.92	1.66	1.99	1.46	1.40	1.49	1.29
PRKCA		Hs00176973_m1	1.26	0.80	1.39	1.56	1.20	1.47	1.11	1.17
PRKCB1		Hs00176998_m1	1.39	1.02	1.05	1.41	1.59	1.19	1.62	2.19
PRKCD		Hs00178914_m1	1.18	0.51	0.99	1.05	1.06	0.94	0.73	0.80
PRKCE		Hs00178455_m1	1.04	0.80	1.35	1.27	1.08	1.57	1.23	1.30
PRKCG		Hs00177010_m1	1.27	1.00	4.22	2.08	1.18	0.93	1.62	0.76
PRKCH		Hs00178933_m1	1.14	0.72	1.15	1.69	1.22	1.16	1.92	1.62
PRKCI		Hs00702254_s1	1.81	0.16	0.15	0.46	0.55	0.81	0.37	0.79
PRKCQ		Hs00234697_m1	1.04	0.69	0.51	1.70	1.08	0.87	1.50	1.67
PRKCZ		Hs00177051_m1	1.21	0.55	0.86	1.10	1.11	0.90	0.84	0.76
PTEN		Hs00829813_s1	1.43	0.19	0.15	0.67	0.31	0.59	0.35	0.80

RQ CD MIII patients											Avrg	SD	CV
8	9	10	11	12	13	14	15	16	17	18			
0.81	1.73	1.12	0.64	0.62	1.10	1.12	1.54	1.17	1.38	0.68	0.98	0.33	34
0.81	1.16	0.90	0.82	1.00	1.16	1.09	1.27	1.40	1.39	1.10	1.06	0.19	18
0.96	1.21	0.84	1.22	1.12	1.18	1.07	0.55	2.13	1.42	0.79	1.13	0.38	34
0.95	1.24	0.50	1.23	1.30	1.06	1.20	1.00	1.30	0.90	1.50	1.13	0.34	30
1.36	1.90	0.90	0.72	0.50	0.75	0.99	1.77	1.19	1.26	0.68	0.94	0.41	43
1.03	3.12	2.50	1.14	1.15	2.01	1.37	0.67	3.31	4.67	1.11	1.88	1.14	61
1.00	1.11	0.97	0.64	0.75	0.91	1.06	1.17	1.24	1.17	0.73	0.91	0.22	24
1.17	1.67	1.14	1.75	2.09	3.14	2.61	1.62	1.92	1.26	1.12	1.52	0.74	48
0.47	0.64	0.63	0.88	0.64	0.83	0.96	1.07	0.77	0.91	0.97	0.83	0.24	29
0.84	1.56	0.50	1.34	1.15	1.96	0.92	1.09	1.27	1.53	1.39	1.20	0.35	29
0.80	1.12	0.87	0.77	0.83	0.81	0.89	1.01	1.28	0.83	0.93	0.91	0.16	18
1.06	0.90	0.59	0.66	0.89	1.51	1.09	1.26	0.99	1.26	0.81	0.92	0.28	30
0.75	1.17	1.13	0.78	1.40	1.78	2.02	1.82	1.35	0.83	1.13	1.16	0.40	34
1.62	3.35	2.37	1.74	1.44	1.92	1.95	1.64	1.09	1.00	1.03	1.60	0.63	40
1.17	1.15	0.83	1.48	1.05	1.68	1.01	0.58	1.73	1.22	1.83	1.24	0.33	27
1.12	1.51	0.83	0.79	0.94	1.11	1.05	0.69	1.25	1.10	0.78	0.97	0.22	23
1.63	1.73	2.29	1.93	1.79	3.40	3.47	1.40	1.17	2.20	2.14	2.14	0.92	43
1.43	1.73	1.00	1.42	1.49	1.42	1.21	0.76	2.48	1.15	1.85	1.47	0.40	27
1.01	2.10	0.86	1.39	1.33	1.16	1.03	0.83	2.08	0.84	1.32	1.30	0.43	33
0.44	0.79	0.68	1.20	1.14	0.98	0.94	0.47	1.12	0.82	1.99	1.07	0.44	41
1.12	1.54	1.02	1.80	1.60	1.54	1.16	1.11	2.53	1.27	1.64	1.45	0.39	27
1.57	2.00	1.10	1.43	1.49	1.70	1.70	1.29	2.33	1.32	1.48	1.53	0.32	21
1.63	3.31	1.48	1.08	1.39	3.07	1.71	1.83	1.83	1.07	0.88	1.48	0.73	50
0.79	0.50	0.28	0.16	0.58	0.92	0.01	0.28	0.87	0.26	0.60	0.52	0.38	73
0.46	0.71	0.45	0.41	0.58	0.50	0.64	0.30	1.16	0.59	0.33	0.56	0.21	37
1.01	1.00	0.82	1.20	0.83	1.25	1.11	1.32	1.06	1.24	0.81	1.05	0.25	23
0.99	1.48	0.96	1.60	1.39	1.86	1.17	0.91	1.65	1.29	2.22	1.43	0.37	26
0.92	1.47	0.59	1.37	1.18	1.01	1.10	0.85	1.64	0.98	1.05	1.16	0.28	24
2.16	2.61	1.43	2.09	2.00	3.53	2.04	2.33	2.24	2.55	1.42	1.92	0.64	33
0.52	1.22	0.96	0.81	1.04	1.10	1.12	0.92	1.26	1.20	1.30	0.97	0.23	24
0.97	1.09	0.95	1.09	0.99	1.54	0.97	1.07	1.30	1.65	1.09	1.18	0.23	20
0.71	1.36	0.62	0.93	2.72	1.50	1.23	1.78	2.97	2.10	1.41	1.62	0.93	57
1.10	2.45	1.21	1.91	1.48	2.29	2.26	1.23	1.91	1.51	1.64	1.58	0.47	30
0.54	1.99	1.30	0.19	0.44	1.35	3.31	1.00	8.00	1.43	13.61	2.03	3.43	169
1.65	2.87	1.26	1.25	1.17	2.00	1.93	1.24	1.86	1.64	1.11	1.44	0.55	38
0.52	1.34	0.81	0.80	0.65	1.27	1.36	1.26	0.96	1.25	0.88	0.96	0.27	28
0.45	2.17	1.59	0.26	0.39	1.02	1.90	0.87	4.54	1.08	6.35	1.32	1.64	125

Symbol	Alias	Assay ID	RQ CD M0							
				1	2	3	4	5	6	7
RAB3B		Hs00267896_m1	1.28	0.99	0.90	0.71	0.75	2.48	2.34	1.24
RAB13		Hs00762784_s1	1.47	0.54	0.93	1.39	1.04	1.17	0.78	1.23
RAC1		Hs00251654_m1	1.41	1.50	2.22	1.83	1.95	3.19	1.76	1.78
RHOA		Hs00357608_m1	1.25	0.90	1.26	1.04	1.41	1.50	1.36	1.03
ROCK1		Hs00178463_m1	1.21	0.82	1.04	0.89	1.17	1.27	1.37	1.25
ROCK2		Hs00178154_m1	1.01	0.67	1.32	1.13	1.05	1.66	1.11	1.02
SEC6L1		Hs00292654_m1	1.03	0.60	1.06	1.55	0.91	0.94	0.96	0.83
SEC8L1		Hs00253986_m1	1.24	0.93	1.35	1.42	1.47	1.25	1.21	1.14
ZAK	MLK7	Hs00370447_m1	0.94	0.85	1.75	1.47	1.11	1.29	0.79	1.31
Actin-binding										
ACTN1		Hs00241650_m1	1.32	0.93	1.18	1.33	1.38	1.66	1.34	1.60
ACTN4		Hs00245168_m1	1.17	0.72	1.46	1.55	1.30	1.63	1.24	0.96
CTTN		Hs00193322_m1	1.60	1.44	1.99	1.86	1.51	1.63	1.54	1.49
EPB41		Hs00270045_m1	1.35	0.85	1.25	1.03	1.45	1.46	1.42	1.27
MLCK		Hs00417542_m1	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.
MYLK		Hs00364926_m1	1.39	0.31	0.49	0.97	0.89	1.79	0.91	1.40
SPTAN1	alpha fodrin	Hs00162203_m1	1.34	1.01	1.19	1.02	1.20	1.46	1.42	0.93
Transcriptional regulators										
ASH1L	huASH1	Hs00218516_m1	1.21	0.73	1.12	1.25	1.28	1.20	1.19	1.28
CDK4		Hs00364847_m1	1.12	1.12	1.45	2.02	1.40	1.15	1.35	1.14
SYMPK		Hs00191361_m1	1.17	0.73	1.33	1.74	1.21	1.25	1.39	1.41
Growth factor										
EGF		Hs00153181_m1	>31.5	>29.9	>31.7	>167	>56.4	>29.5	>76.1	b.d.
EGFR		Hs00193306_m1	1.24	0.81	1.25	1.08	1.33	1.36	0.99	1.02

Abbreviations: RQ, relative quantification; Avrg, average; SD, standard deviation; CV, coefficient of variation (SD/Avrg x 100); b.d., below detection. Upward (Avrg – SD > 1) and downward (Avrg + SD < 1) trends are shaded. Boxed numbers indicate outlier-values in the range.

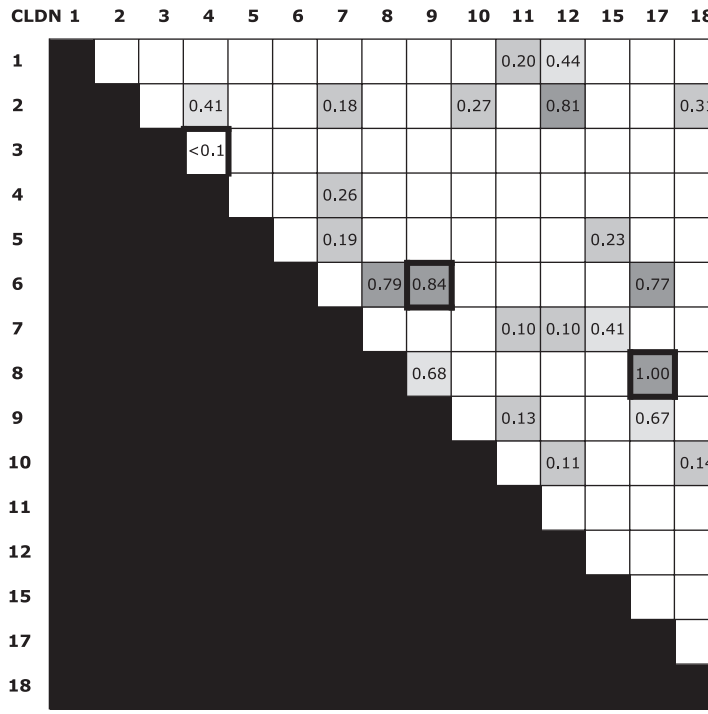
but have different genomic locations. The co-regulated claudins also showed, as may be expected, similar expression levels in the normal duodenum (figure 2A).

Compared to the claudins, the expression variability for the other TJ genes was rather limited (see CV-values in table 2). The changes in gene expression, if present, were also rather subtle and could not be marked as statistically significant and we would rather refer to them as upward or downward trends. The design of our study (comparison of pooled controls with individual patients) also precluded application of an appropriate statistical test since we have no information on the level of variation within the control group. In table 2 we marked all genes with upward (average – SD > 1) and downward (average + SD < 1) trends in expression.

RQ CD MIII patients											Avrg	SD	CV
8	9	10	11	12	13	14	15	16	17	18			
1.45	0.83	0.86	0.78	1.35	1.38	1.05	0.85	3.09	0.63	1.65	1.29	0.69	54
0.54	1.16	0.67	0.79	0.89	1.10	1.15	0.52	2.91	1.05	3.87	1.21	0.85	70
1.37	1.77	1.62	1.51	2.63	2.13	1.75	0.75	3.52	1.00	1.40	1.87	0.69	37
1.47	1.89	1.33	1.12	1.27	1.50	1.64	1.27	1.97	1.27	1.66	1.38	0.28	21
1.91	2.12	1.08	1.33	1.26	1.54	1.31	1.04	2.07	1.30	1.11	1.33	0.37	28
1.00	1.12	0.87	0.95	1.18	1.16	0.94	0.97	1.68	1.00	1.27	1.12	0.25	22
0.70	0.92	0.85	0.77	0.83	1.20	1.28	1.32	0.99	1.05	0.74	0.97	0.24	25
1.33	1.58	1.39	1.16	1.42	2.16	2.26	1.82	1.37	1.82	1.00	1.45	0.36	25
1.32	0.82	0.33	1.11	0.94	1.03	0.58	0.42	1.90	0.72	1.59	1.07	0.44	41
1.36	1.38	0.97	1.36	1.82	2.30	1.64	1.27	2.02	1.31	1.57	1.47	0.34	23
0.94	1.41	0.78	1.25	1.18	1.20	1.22	1.06	1.61	1.20	1.22	1.22	0.26	21
1.30	1.94	1.30	1.39	1.50	1.38	2.02	1.28	2.49	1.28	1.88	1.62	0.34	21
1.56	2.20	1.28	1.26	1.31	1.75	1.47	1.18	1.80	1.24	1.42	1.40	0.30	22
b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	>12.8			
0.93	1.60	0.45	0.76	1.32	1.73	1.04	0.62	2.08	0.90	0.79	1.05	0.50	47
0.95	1.19	1.14	1.28	1.17	1.52	1.34	0.94	1.64	1.10	1.42	1.22	0.21	17
1.13	1.69	1.21	1.08	1.29	1.77	1.49	1.23	1.69	1.70	1.10	1.30	0.27	21
0.96	0.97	1.17	1.29	1.34	1.73	1.81	1.46	1.67	0.88	1.72	1.37	0.32	24
0.97	1.02	0.90	1.40	1.49	1.68	1.43	1.23	1.20	1.30	1.40	1.28	0.26	20
>29.0	>34.0	>51.8	>116	b.d.	>85.8	b.d.	b.d.	>49.1	>33.6	>70.5	>59.5	39.3	66
0.99	1.07	0.66	1.00	0.99	1.31	1.48	1.13	1.23	1.20	1.14	1.11	0.20	18

Of the adaptors only *CGN* and MAG1 showed an upward, and *INADL (PATJ)* a downward trend. A quarter of the TJ genes involved in signal transduction showed trends of differential expression (ten upwards, two downwards), including six members from the phosphatase *PPP2* family, two protein kinase C genes, and the two small GTPases *RAC1* and *RHOA*. Furthermore, four out of the six expressed genes coding for actin-binding proteins showed an upward trend, similar to the three transcriptional activators. The growth factor *EGF*, which was not detectable in the normal duodenum, was strongly induced (average > 60-fold) in 14 out of the 18 MIII cases and also highly increased (average > 30-fold) in M0 remission patients. The receptor *EGFR*, however, remained unchanged.

A



DISCUSSION

With this study we present the first comprehensive analysis of the transcriptional activity of genes comprising the TJ network. We examined this pathway in the duodenal mucosa of controls and CD patients. The pathology of CD is characterized by enhanced permeability suggesting an altered TJ function[3]. The transmembrane proteins occludin, JAM, and the claudin family are the major structural components of the TJ that facilitate barrier function and selective paracellular transport[9]. Members of the claudin familie (currently 24) form a complex labyrinth of intercellular strands that acts as a dynamic molecular sieve[15, 16]. Changes in the composition of claudins in the TJ are accompanied by an altered transepithelial flux[17]. The claudin expression profile that we observed in the duodenum is dominated by *CLDN15*, *-3*, and *–7*, and to lesser extent by *CLDN12*, *-4*, *-2*, and *–17*. Remarkably, this claudin transcriptional profile is evolutionary conserved in rodents[18-20], indicating that this tight regulatory control emerged over 75 millions years ago. The small intestinal claudin profile is also distinct from that in other sections of the gastrointestinal tract, as evident from studies in mice[20] and man[17],

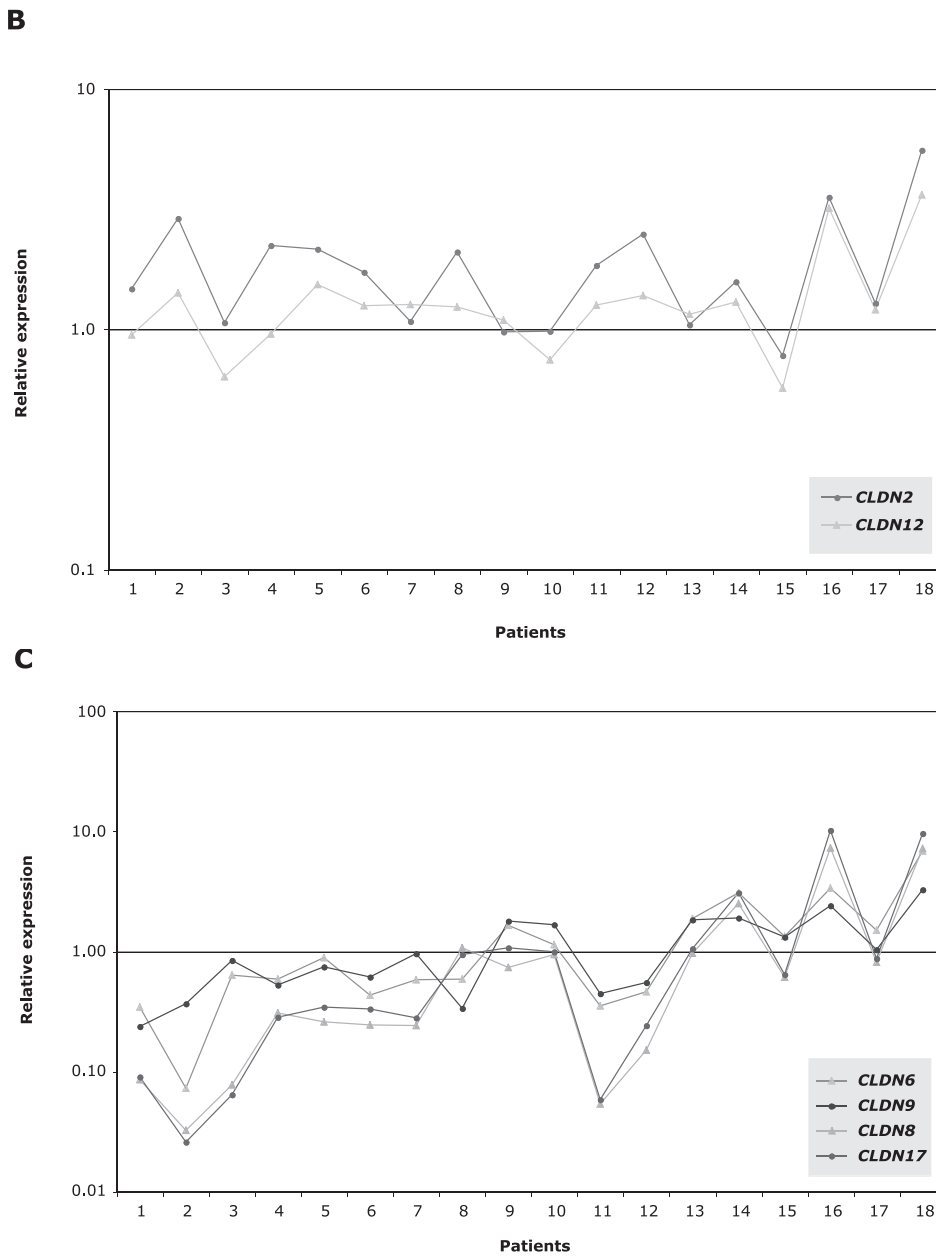


Figure 3. Co-regulation of claudin genes in the duodenal mucosa of atrophic CD MIII patients. (**A**) Pair-wise comparison of claudin expression values derived from 18 CD MIII cases. Standard correlations are indicated and color marked (orange, $r^2 > 0.7$; yellow, $0.4 > r^2 > 0.7$; green, $0.1 > r^2 > 0.4$; blanc, $r^2 < 0.1$). Heavy-lined squares indicate pairs of claudins that are juxtaposed in the genome. The five highly correlating claudin pairs could be grouped in two expression profiles consisting of one (**B**), and two (**C**) matching claudin pairs.

and reflects the different permeability requirements along the digestive system. We also observed strong co-regulation between three pairs of claudins of which two pairs (*CLDN6* and *–9*; *CLDN8* and *-17*) are juxtaposed in the genome and are likely controlled by shared local sequence elements. The third pair (*CLDN2* and *-12*) was located on different chromosomes and their co-expression is possibly imposed by functional constraints. Co-regulation of these pairs of claudins suggests that together they may have similar effects on the permeability. Compared to the other TJ genes the claudins were marked by extreme fluctuations in expression between individual CD MIII cases, and 15 out of the 18 patients had at least one outlier-value. Particularly, *CLDN8*, *-10*, *-17*, and *–18* contributed to this variability. This could result in distorted claudin profiles with consequences for the permeability, although we cannot completely rule out the contribution of differences in the genetic backgound independent of the pathology.

From the adaptors only *CGN*, *INADL* (*PATJ*), and *MAGI11* showed suggestive differential expression. The ZO-1 protein, coded for by *TJP1*, is commonly used as a marker for a distorted TJ function[3, 11]. However, *TJP1* showed no transcriptional change in patients underscoring its regulation by post-translational processes and protein trafficking[9]. Most of the upward expression trends in MIII patients were in the groups of genes involved in signal transduction, actin-binding, and transcriptional regulation. The observed changes in the transcriptional activity of protein kinases (*PKC*) and phosphatases (*PPP2*) may have a joint effect on TJ function, cell polarity, and the cytoskeleton[8]. Changes in F-actin are associated with TJ impairment[21]. Both, the actin-binding genes (particularly *ACTN4*) and the major regulator of the actinomyosin cytoskeleton, *RHOA*, showed high expression with an upward trend in CD patients. Recently, we reported genetic association between CD and *MYO9B*, an atypical myosin that carries a RhoGAP domain[22]. Both, the pathology and the genetics may augment RhoA protein expression, resulting in altered permeability and enhanced influx of gluten peptides, and thus further exacerbate the immune response. Finally, the dramatic increase in *EFG* expression, known for its epithelium and permeability preserving effect[23, 24], may be the intestine's natural response to cope with the potentially dangerous barrier defect.

The TJ transcriptome of the normal and CD-affected duodenum that we presented here lays a comprehensive foundation for further expression research in CD and other intestinal disorders, whether by using gene expression or proteomics, to further unravel the intricate fabric of permeability, polarity, and differentiation.

ACKNOWLEDGEMENTS

We like to thank Drs. G.A. Meijer (VUmc, Amsterdam) and J.W.R. Meijer (Rijnstate, Arnhem) for Marsh typing, and Drs. M.L. Mearin and J. Schweizer

(LUMC, Leiden) for additional patient samples. We are indebted to the patients for their cooperation. Thanks to Jackie Senior for editing the manuscript. This study was financially supported by the Netherlands Organization for Scientific Research (NWO, VICI-grant 918.66.620), and the Celiac Disease Consortium, an Innovative Cluster approved by the Netherlands Genomics Initiative and partially funded by the Dutch Government (BSIK03009).

REFERENCES

1. Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. Nat Rev Immunol 2002;2:647-55.
2. Green PH, Jabri B. Coeliac disease. Lancet 2003;362:383-91.
3. Drago S, El Asmar R, Di Pierro M, et al. Gliadin, zonulin and gut permeability: Effects on coeliac and non-coeliac intestinal mucosa and intestinal cell lines. Scand J Gastroenterol 2006;41:408-19.
4. Clemente MG, De Virgiliis S, Kang JS, et al. Early effects of gliadin on enterocyte intracellular signalling involved in intestinal barrier function. Gut 2003;52:218-23.
5. Mallant-Hent R, Mary B, von Blomberg E, et al. Disappearance of anti-Saccharomyces cerevisiae antibodies in coeliac disease during a gluten-free diet. European journal of gastroenterology & hepatology 2006;18:75-8.
6. Granito A, Muratori L, Muratori P, et al. Anti-saccharomyces cerevisiae antibodies (ASCA) in coeliac disease. Gut 2006;55:296.
7. Sawada N, Murata M, Kikuchi K, et al. Tight junctions and human diseases. Med Electron Microsc 2003;36:147-56.
8. Matter K, Balda MS. Signalling to and from tight junctions. Nat Rev Genet 2003;4:225-36.
9. Schneeberger EE, Lynch RD. The tight junction: a multifunctional complex. Am J Physiol Gastrointest Liver Physiol 2004;286:C1213-28.
10. Matter K, Aijaz S, Tsapara A, et al. Mammalian tight junctions in the regulation of epithelial differentiation and proliferation. Curr Opin Cell Biol 2005;17:453-8.
11. Ciccocioppo R, Finamore A, Ara C, et al. Altered expression, localization, and phosphorylation of epithelial junctional proteins in coeliac disease. Am J Clin Pathol 2006;125:502-11.
12. Report of a working group of the United European Gastroenterology Week in Amsterdam. When is a coeliac a coeliac? Eur J Gastroenterol Hepatol 2001;13:1123-28.
13. Wapenaar MC, van Belzen MJ, Fransen JH, et al. The interferon gamma gene in coeliac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility. J Autoimmun 2004;23:183-90.
14. Lapierre LA, Tuma PL, Navarre J, et al. VAP-33 localizes to both an intracellular vesicle population and with occludin at the tight junction. J Cell Sci 1999;112 (Pt 21):3723-32.
15. Sasaki H, Matsui C, Furuse K, et al. Dynamic behavior of paired claudin strands within apposing plasma membranes. Proceedings of the National Academy of Sciences of the United States of America 2003;100:3971-6.

16. Shen L, Turner JR. Role of epithelial cells in initiation and propagation of intestinal inflammation. Eliminating the static: tight junction dynamics exposed. Am J Physiol Gastrointest Liver Physiol 2006;290:G577-82.
17. Van Itallie CM, Anderson JM. Claudins and epithelial paracellular transport. Annu Rev Physiol 2006;68:403-29.
18. Rahner C, Mitic LL, Anderson JM. Heterogeneity in expression and subcellular localization of claudins 2, 3, 4, and 5 in the rat liver, pancreas, and gut. Gastroenterology 2001;120:411-22.
19. Inai T, Sengoku A, Guan X, et al. Heterogeneity in expression and subcellular localization of tight junction proteins, claudin-10 and -15, examined by RT-PCR and immunofluorescence microscopy. Arch Histol Cytol 2005;68:349-60.
20. Fujita H, Chiba H, Yokozaki H, et al. Differential expression and subcellular localization of claudin-7, -8, -12, -13, and -15 along the mouse intestine. J Histochem Cytochem 2006;54:933-44.
21. Bruewer M, Hopkins AM, Hobert ME, et al. RhoA, Rac1, and Cdc42 exert distinct effects on epithelial barrier via selective structural and biochemical modulation of junctional proteins and F-actin. Am J Physiol Gastrointest Liver Physiol 2004;287: C327-35.
22. Monsuur AJ, de Bakker PI, Alizadeh BZ, et al. Myosin IXB variant increases the risk of coeliac disease and points toward a primary intestinal barrier defect. Nature genetics 2005;37:1341-4.
23. Sheng G, Bernabe KQ, Guo J, et al. Epidermal growth factor receptor-mediated proliferation of enterocytes requires p21waf1/cip1 expression. Gastroenterology 2006;131:153-64.
24. Basuroy S, Seth A, Elias B, et al. MAPK interacts with occludin and mediates EGF-induced prevention of tight junction disruption by hydrogen peroxide. Biochem J 2006;393:69-77.

Chapter 7

Association of tight junction genes *UMCU23* and *UMCU29* with gluten sensitive enteropathy and inflammatory bowel disease implies a common barrier defect

MC Wapenaar, AJ Monsuur, K Hunt, P Howdle, G Holmes,
DA van Heel, AA van Bodegraven, CJJ Mulder, and C Wijmenga

ABSTRACT

Coeliac disease, or gluten-sensitive enteropathy (GSE), and inflammatory bowel disease (IBD) are common gastrointestinal disorders. Both diseases display enhanced intestinal permeability, thought to emanate from gluten exposure (GSE) or bacterial overgrowth (IBD). Previous studies showed association of both conditions to variants in the *MYO9B* gene, suggesting to be involved in enhanced epithelial permeability. We hypothesized that gene variants in tight junction genes might further influence epithelial barrier function, thus contributing to the cause of both GSE and IBD. This hypothesis was tested by analysis of 41 tight junction genes with 201 tag SNPs. Two genes, *UMCU29* and *UMCU23*, showed association with GSE in a Dutch cohort. Replication in a British GSE cohort and joint analysis further substantiated the data for both *UMCU29* ($P = 6x10^{-5}$; OR 1.24, 95% CI 1.12-1.38) and *UMCU23* ($P = 0.0006$; OR 1.2, 95% CI 1.08-1.33). Association was also observed in Dutch patients between *UMCU29* and ulcerative colitis (P = 0.0043; OR 1.3, 95% CI 1.07-1.59), and *UMCU23* and IBD (P = 0.0047; OR 1.22, 95% CI 1.05-1.41). This suggests that GSE and IBD share a common etiology through tight junction-mediated gut permeability.

RESULTS AND DISCUSSION

Gluten-sensitive enteropathy (GSE) and inflammatory bowel disease (IBD) are two common gastrointestinal inflammatory disorders both characterized by an enhanced intestinal epithelial permeability[1, 2]. A proportion of first-degree relatives of IBD patients also display impaired barrier function, suggesting that this is heritable rather than acquired[3]. In further support of this, it has been shown in IBD in human[4] and mouse[5], and GSE in human[6] and dog[7], that an impaired intestinal permeability is present long before the onset of disease. Moreover, treatment of GSE patients on a gluten-free diet reverses the disease process but does not completely restore the increased intestinal permeability[8]. Both GSE and IBD have a strong genetic component and there are currently tens of susceptibility loci in the human genome linked to either IBD[9] or GSE[10]. Notably, these studies show sharing of certain chromosomal regions that may predispose to both disorders, e.g. a locus on 5q31-q33 and a locus on 19p13[11]. GSE and IBD co-occur in families and patients, with an approximately five-fold increased prevalence of IBD in GSE patients[12-15]. Together, this suggests that IBD and GSE share part of their genetic susceptibility at the molecular level. Recently, the GSE-associated gene *MYO9B*[16], which localizes to the *IBD6* locus on 19p13, was shown to be strongly associated with ulcerative colitis and to a lesser extend to Crohn's disease (together comprising the IBD phenotype) in four different European populations, i.e. Dutch, British, Italian, and French-Canadian[17]. It was hypothesized that myosin IXB – which contains a Rho-GTPase activating domain (GAP) – is involved in intestinal permeability through

remodelling of the cytoskeleton and tight junction assembly, through its interaction with RhoA[16, 18]. This might suggest that the barrier defect observed in GSE and IBD is mediated through epithelial tight junctions. Functional abnormalities of tight junctions have been observed in non-inflamed ileum of Crohn's disease as well, suggesting an implication in the pathogenesis of IBD[19]. Similarly, altered expression, localization, and phosphorylation of epithelial junctional proteins have been observed in GSE[20].

Based on the above, genes encoding tight junction proteins should be considered highly relevant functional candidate genes for GSE and IBD. In the present study we focus on 41 genes from the tight junction pathway, including those encoding for transmembrane, adaptor, signal transduction, and transcriptional-regulatory proteins[21, 22]; (KEGG: http://www.genome.jp/kegg/pathway.html). In a first screen 201 tag SNPs in these genes were successfully genotyped in a cohort of 463 Dutch GSE cases and 470 Dutch controls (see Supplementary Table 1 online). We observed evidence in favor of association ($P_{uncorrected} < 0.01$) for five SNPs in two genes (*UMCU23* and *UMCU29*) (Table 1). UMCU23 and UMCU29 are both tight junction adaptor proteins that act as membrane-associated scaffolds. Given that these 41 tight junction genes were selected based on prior knowledge, it is not clear what level of proof one would require accepting association. If one would correct for testing 41 functional candidate genes, SNP rsUMCU117 in *UMCU23* would remain significant (Pc < 0.05/41 = 0.0012). A follow-up study of the five SNPs with P < 0.01 was performed by typing 459 extra Dutch controls, and a second, fully independent, cohort of 658 unrelated British GSE patients and 1185 British controls. For all five SNPs allele frequencies between the three different control cohorts (two Dutch, one British) did not differ significantly. Association analysis of the 463 Dutch GSE cases versus 929 Dutch controls improved the P values for two of the SNPs, one in each gene (Table 1). The three SNPs in *UMCU23* are not in linkage disequilibrium with each other (D'<0.06, r2<0.003) and only weak correlation is observed between the two SNPs in *UMCU29* (D'=0.85, r2=0.40). Although the change in allele frequencies in the British GSE cases was similar but less pronounced compared to the Dutch GSE cases, only SNP rsUMCU172 in *UMCU29* showed positive association (P = 0.044). Combining the Dutch and British cohorts (1121 cases versus 2114 controls) strengthened the association considerably and revealed highly significant association for three of the five SNPs (Table 1). We observed the smallest P value for rsUMCU172 (P = 6 x 10-5) located in intron 21 of *UMCU29*. Individuals carrying the A allele have a modest but significantly higher risk of developing GSE (OR 1.24; 95% CI 1.12 – 1.38). The most significant SNP in *UMCU23* was rsUMCU117 (P = 0.0006) located in intron 14; this SNP was associated with a 1.2-fold increased risk for GSE (OR 1.2; 95% CI 1.08 – 1.33). As both these associated SNPs are in

the introns and are not expected to have any functional effect, these SNPs can be considered markers of disease susceptibility.

Table 1. Association analysis of *UMCU29* and *UMCU23* SNPs in gluten-sensitive enteropathy

		GSE cases			Controls			P by cohort*
		Allele counts			Allele counts			
		Major	Minor	MAF	Major	Minor	MAF	
UMCU29	rsUMCU172							
Dutch		417	439	51.3	510	402	44.1	0.0024
Dutch extra controls		417	439	51.3	1014	782	43.5	0.0001
UK		627	653	51.0	1221	1105	47.5	0.022
Pooled Dutch + UK		1044	1092	51.1	2235	1887	45.8	0.00003
UMCU29	rsUMCU173							
Dutch		538	316	37.0	636	276	30.3	0.0027
Dutch extra controls		538	316	37.0	1233	561	31.3	0.0017
UK		787	489	38.3	1496	814	35.2	0.033
Pooled Dutch + UK		1325	805	37.8	2729	1375	33.5	0.0004
UMCU23	rsUMCU117							
Dutch		397	459	53.6	494	418	45.8	0.0011
Dutch extra controls		397	459	53.6	960	840	46.7	0.0004
UK		647	635	49.5	1235	1077	46.6	0.045
Pooled Dutch + UK		1044	1094	51.2	2195	1917	46.6	0.0003
UMCU23	rsUMCU129							
Dutch		484	372	43.5	574	338	37.1	0.0061
Dutch extra controls		484	372	43.5	1115	685	38.1	0.0038
UK		805	489	37.8	1421	883	38.3	0.38
Pooled Dutch + UK		1289	861	40.0	2536	1568	38.2	0.08
UMCU23	rsUMCU140							
Dutch		464	392	45.8	550	362	39.7	0.0095
Dutch extra controls		464	392	45.8	1063	739	41.0	0.0099
UK		542	362	40.0	779	539	40.9	0.34
Pooled Dutch + UK		1006	754	42.8	1842	1278	41.0	0.10

*The initial analysis of 483 Dutch GSE cases and 470 Dutch controls are two-tailed P values calculated using Haploview. On the basis of these findings we determined a priori the A variant of rsUMCU172, the A variant of rsUMCU173, the G variant of rsUMCU117, the A variant of rsUMCU129, and the A variant of rsUMCU140 as the reference alleles for further analysis. This allowed us to perform one-sided hypothesis testing for the follow-up cohorts (i.e. Dutch extra controls, UK and pooled data).

Next we were interested to learn if these two tight junction adaptor proteins were also associated to IBD and its sub-phenotypes Crohn's disease and ulcerative colitis. In total 298 Dutch Crohn's disease and 290 Dutch ulcerative colitis cases were genotyped with the same five tag SNPs in *UMCU23* and *UMCU29*. The entire IBD population of 588 cases showed significant association with rsUMCU117 in *UMCU23* (P = 0.0094; OR 1.22, 95% CI 1.05-1.41); both Crohn's disease and

ulcerative colitis contributed to this association (Table 2). Ulcerative colitis also showed significant association to rsUMCU173 in *UMCU29* (P = 0.0085; OR 1.3, 95% CI 1.07-1.59) whereas this SNP did not show association in the entire IBD cohort (P = 0.085) and in Crohn's disease (P = 0.92).

Table 2. Association analysis of *UMCU29* and *UMCU23* SNPs in inflammatory bowel disease

		IBD cases			Controls			P by cohort*
		Allele counts			Allele counts			
		Major	Minor	MAF	Major	Minor	MAF	
UMCU29	rsUMCU172							
IBD		642	496	43.6	1014	782	43.5	0.49
Crohn's disease		322	256	44.3	1014	782	43.5	0.38
Ulcerative colitis		320	240	42.9	1014	782	43.5	0.39
UMCU29	rsUMCU173							
IBD		750	392	34.3	1233	561	31.3	0.042
Crohn's disease		396	182	31.5	1233	561	31.3	0.46
Ulcerative colitis		354	210	37.2	1233	561	31.3	0.0043
UMCU23	rsUMCU117							
IBD		550	586	51.6	960	840	46.7	0.0047
Crohn's disease		277	297	51.7	960	840	46.7	0.017
Ulcerative colitis		273	289	51.4	960	840	46.7	0.024
UMCU23	rsUMCU129							
IBD		705	431	37.9	1115	685	38.1	0.48
Crohn's disease		355	219	38.2	1115	685	38.1	0.48
Ulcerative colitis		350	212	37.7	1115	685	38.1	0.44
UMCU23	rsUMCU140							
IBD		670	452	40.3	1063	739	41.0	0.35
Crohn's disease		339	231	40.5	1063	739	41.0	0.42
Ulcerative colitis		331	221	40.0	1063	739	41.0	0.34

*The initial analysis in GSE was based on two-sided analysis using Haploview. On the basis of these findings we determined a priori the A variant of rsUMCU172, the A variant of rsUMCU173, the G variant of rsUMCU117, the A variant of rsUMCU129, and the A variant of rsUMCU140 as the reference alleles for further analysis. This allowed us to perform one-sided hypothesis testing for the follow-up cohorts.

Weakening of the barrier function of tight junctions can result from natural manipulation by pathogens[23, 24], as well as by food substances such as gluten[25, 26]. This study shows for the first time that also variants in tight junction adaptor proteins might be the cause of various gastrointestinal disorders such as GSE and IBD. The two associated genes encode for the UMCU29 and UMCU23 adaptor proteins involved in tight junction assembly. Several of the membrane-associated proteins of the tight junction complex interact with the actin cytoskeleton, and signalling to the actin cytoskeleton is important in regulating both tight junction assembly and function, with a crucial role for RhoA-GTPase activity[27, 28]. Myosin

IXB might also have a role in this process, as it possesses a Rho-GAP domain that can negatively regulate Rho proteins. Since primary barrier defects are expected to be subtle, perpetuation of the cycle of barrier malfunction is presumably sustained by the subsequent local inflammation as it is well established that proinflammatory cytokines - such as IFN-γ and TNF-α - further trigger barrier dysfunction and bacterial translocation.

Both GSE and ulcerative colitis are characterized by inflammation restricted to the intestinal mucosa. In this respect it is interesting to note that both *MYO9B* and *UMCU29* were predominantly associated with ulcerative colitis. A healthy gut mucosa is characterized by a robust and selective barrier, maintained by properly differentiated and polarized epithelial cells. Weakening of the barrier due to external factors like gluten or intestinal bacteria does not only result in an inflammation, but also leads to a loss of differentiation and polarization, which results in clinical syndromes characterized by nutrient malabsorption and diarrhea. The outcome of the interaction between potentially harmful external factors and the mucosal barrier may be influenced by genetic variants in the genes that code for the tight junction complex.

In conclusion, the association of *MYO9B, UMCU29* and *UMCU23* in both GSE and IBD suggests a common defect in tight junction mediated epithelial barrier function. The recent identification of *DLG5* as a susceptibility gene for IBD is in line with our observations since DLG5 is also involved in epithelial cell-cell contacts and polarity[29].

METHODS

Patients and controls

The cohorts are summarized in Supplementary table 2. DNA, isolated from whole blood, was available from a cohort of 463 Dutch coeliac disease patients[16] and 658 British coeliac disease patients[30]. The 588 Dutch IBD cases are described in detail elsewhere[31]. There were two cohorts of random Dutch blood bank donors available: Set 1 controls (n = 470, described by Monsuur et al., 2005)[16] and set 2 controls (n = 459). The 1185 British controls were drawn from the 1958 British Birth Cohort. All Dutch cases and controls were from the Netherlands, of European descent, and with at least 3 of the 4 grandparents also born in the Netherlands. This study was approved by the Medical Ethical Committee of the University Medical Center Utrecht, the VU University Medical Center Amsterdam Ethics Committee, and SouthEast Multicentre Research Ethics Committee, Oxford Research Ethics Committee, Guy's Hospital Research Ethics Committee, Harrow Research Ethics Committee.

Comprehensive screen: tag SNP selection and genotyping

We selected 44 tight junction genes based on a tight junction network in the literature[18, 21, 22] (KEGG: www.genome.jp/kegg/). Three genes were excluded for further study: for *UMCU42* and *UMCU43* no tag SNPs could be selected, and *UMCU44* was localized on the X-chromosome. SNPs were selected by downloading all the SNPs typed in the CEPH (Utah residents with ancestry from northern and western Europe) population located in the genomic sequence of the remaining 41 genes from the HapMap database[32] (November 2004, Phase I, http://www.hapmap.org/). From these SNPs, the program Tagger (available at http://www.broad.mit.edu/mpg/tagger/) was used to select tag SNPs so that all SNPs with a minor allele frequency (MAF) ≥ 10% was captured with $r^2 \geq 0.7$ (excluding SNPs with low Illumina quality design scores). For some genes the amount of tag SNPs was too large and there we tagged all exons with exon-intron boundaries and used a MAF ≥ 20% (see Supplementary Table 3). A final set of 215 tag SNPs was obtained for genotype analysis in 463 Dutch coeliac cases and set 1 controls (n=470). See Supplementary Table 4 online for the detailed list of SNPs.

SNP genotyping was performed using the GoldenGate assay on an Illumina BeadStation 500 GX (Illumina Inc., San Diego, USA). All tag SNPs were examined for their resulting quality and 13 who had a low signal or too wide clusters were excluded. In addition, one tag SNP was not in Hardy-Weinberg equilibrium (HWE) in the controls, yielding a total of 201 SNPs that were successfully analyzed (93%).

The five SNPs that revealed $P < 0.01$ in the comprehensive screen were genotyped as Assays on Demand (Aplied Biosystems) in the set 2 Dutch controls (n=459), the British coeliac cases (n=658) and controls (n=1185) and the Dutch IBD cases (n=588). (*Remark: In case of SNP rsUMCU140, temporarily only 470 British coeliac cases and 657 controls could be tested because of practical reasons*). No allele frequency differences were observed between the Dutch control cohorts (See Supplementary Table 5).

Statistical analysis

Association chi-squares and two-tailed *P* values were calculated using the Haploview program (freely available at http://www.hapmap.org), for each stage of the study. SNPs that were not in HWE ($P \leq 0.001$) in the controls were excluded for further analysis. We used multiple logistic regression analysis to estimate allelic and genotypic odds ratios (OR) and the corresponding 95% confidence intervals for the five SNP tested in the follow-up studies. All the analyses were performed using STATA statistical software, version 8.0 for MS Windows.

ACKNOWLEDGEMENTS

We thank all the patients, their physicians, and the Dutch Coeliac Disease Foundation for participating in this study. We thank Harry van Someren as database manager, and Jackie Senior for critically reading the manuscript. We acknowledge use of DNA from the British 1958 Birth Cohort collection, funded by the UK Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02, and the help of Prof David Strachan (St Georges University of London) in acquiring these samples. We also thank Prof Raymond J Playford and Prof Parveen Kumar (Barts and the London), Prof Derek P Jewell and Dr Simon PL Travis (John Radcliffe Hospital, Oxford), and Dr Julian RF Walters (Hammersmith Hospital, London). This study was supported by grants from the Netherlands Organization for Scientific Research (grants 902-22-094 and 912-02-028), the Dutch Digestive Disease Foundation (grants 97-44 and 03-06) and the Coeliac Disease Consortium, an innovative cluster approved by the Netherlands Genomics Initiative and partially funded by the Dutch government (grant BSIK03009).

COMPETING INTEREST STATEMENT

The authors declare they have no competing financial interests.

REFERENCES

1. Fasano A, Shea-Donohue T. Mechanisms of disease: the role of intestinal barrier function in the pathogenesis of gastrointestinal autoimmune diseases. Nat Clin Pract Gastroenterol Hepatol 2005;2:416-22.
2. Schreiber S, Rosenstiel P, Albrecht M, et al. Genetics of Crohn disease, an archetypal inflammatory barrier disease. Nat Rev Genet 2005;6:376-88.
3. Halme L, Paavola-Sakki P, Turunen U, et al. Family and twin studies in inflammatory bowel disease. World J Gastroenterol 2006;12:3668-72.
4. Ma TY. Intestinal epithelial barrier dysfunction in Crohn's disease. Proc Soc Exp Biol Med 1997;214:318-27.
5. Sugawara K, Olson TS, Moskaluk CA, et al. Linkage to peroxisome proliferator-activated receptor-gamma in SAMP1/YitFc mice and in human Crohn's disease. Gastroenterology 2005;128:351-60.
6. Schulzke JD, Schulzke I, Fromm M, et al. Epithelial barrier and ion transport in coeliac sprue: electrical measurements on intestinal aspiration biopsy specimens. Gut 1995;37:777-82.
7. Hall EJ, Batt RM. Abnormal permeability precedes the development of a gluten sensitive enteropathy in Irish setter dogs. Gut 1991;32:749-53.
8. Smecuol E, Sugai E, Niveloni S, et al. Permeability, zonulin production, and enteropathy in dermatitis herpetiformis. Clin Gastroenterol Hepatol 2005;3:335-41.
9. Newman B, Siminovitch KA. Recent advances in the genetics of inflammatory bowel disease. Curr Opin Gastroenterol 2005;21:401-7.

10. van Heel DA, Hunt K, Greco L, et al. Genetics in coeliac disease. Best Pract Res Clin Gastroenterol 2005;19:323-39.
11. Curley CR, Monsuur AJ, Wapenaar MC, et al. A functional candidate screen for coeliac disease genes. 2006.
12. Cottone M, Marrone C, Casa A, et al. Familial occurrence of inflammatory bowel disease in coeliac disease. Inflamm Bowel Dis 2003;9:321-3.
13. Schedel J, Rockmann F, Bongartz T, et al. Association of Crohn's disease and latent coeliac disease: a case report and review of the literature. Int J Colorectal Dis 2005;20:376-80.
14. Tursi A, Giorgetti GM, Brandimarte G, et al. High prevalence of coeliac disease among patients affected by Crohn's disease. Inflamm Bowel Dis 2005;11:662-6.
15. Yang A, Chen Y, Scherl E, et al. Inflammatory bowel disease in patients with coeliac disease. Inflamm Bowel Dis 2005;11:528-32.
16. Monsuur AJ, de Bakker PI, Alizadeh BZ, et al. Myosin IXB variant increases the risk of coeliac disease and points toward a primary intestinal barrier defect. Nature genetics 2005;37:1341-4.
17. Van Bodegraven AA, Curley CR, Hunt K, et al. Genetic variation in myosin IXB is associated with ulcerative colitis. Gastroenterology 2006;in press.
18. Matter K, Aijaz S, Tsapara A, et al. Mammalian tight junctions in the regulation of epithelial differentiation and proliferation. Curr Opin Cell Biol 2005;17:453-8.
19. Soderholm JD, Olaison G, Peterson KH, et al. Augmented increase in tight junction permeability by luminal stimuli in the non-inflamed ileum of Crohn's disease. Gut 2002;50:307-13.
20. Ciccocioppo R, Finamore A, Ara C, et al. Altered expression, localization, and phosphorylation of epithelial junctional proteins in coeliac disease. Am J Clin Pathol 2006;125:502-11.
21. Schneeberger EE, Lynch RD. The tight junction: a multifunctional complex. Am J Physiol Gastrointest Liver Physiol 2004;286:C1213-28.
22. Matter K, Balda MS. Signalling to and from tight junctions. Nat Rev Genet 2003;4:225-36.
23. Berkes J, Viswanathan VK, Savkovic SD, et al. Intestinal epithelial responses to enteric pathogens: effects on the tight junction barrier, ion transport, and inflammation. Gut 2003;52:439-51.
24. Clayburgh DR, Shen L, Turner JR. A porous defense: the leaky epithelial barrier in intestinal disease. Lab Invest 2004;84:282-91.
25. Sander GR, Cummins AG, Henshall T, et al. Rapid disruption of intestinal barrier function by gliadin involves altered expression of apical junctional proteins. FEBS Lett 2005;579:4851-5.
26. Drago S, El Asmar R, Di Pierro M, et al. Gliadin, zonulin and gut permeability: Effects on coeliac and non-coeliac intestinal mucosa and intestinal cell lines. Scand J Gastroenterol 2006;41:408-19.

27. Takakuwa R, Kokai Y, Kojima T, et al. Uncoupling of gate and fence functions of MDCK cells by the actin-depolymerizing reagent mycalolide B. Exp Cell Res 2000;257:238-44.
28. Rajasekaran SA, Palmer LG, Moon SY, et al. Na,K-ATPase activity is required for formation of tight junctions, desmosomes, and induction of polarity in epithelial cells. Mol Biol Cell 2001;12:3717-32.
29. Stoll M, Corneliussen B, Costello CM, et al. Genetic variation in DLG5 is associated with inflammatory bowel disease. Nature genetics 2004;36:476-80.
30. Hunt KA, Monsuur AJ, McArdle W, et al. Lack of association of MYO9B genetic variants with coeliac disease in a British cohort. 2006.
31. Xia B, Crusius JB, Wu J, et al. Signal transducer and activator of transcription 6 gene G2964A polymorphism and inflammatory bowel disease. Clinical and experimental immunology 2003;131:446-50.
32. The International HapMap Consortium. The International HapMap Project. Nature 2003;426:789-96.

Supplementary table 2. Demographics of the cohorts

Cohort	Number	Age (range)	% female
Dutch GSE cases[a]	463	44 (3-93)	69.1
Dutch controls (Set 1)[a]	470	48 (19-70)	41.7
Dutch controls (Set 2)	459	49 (20-74)	35.7
English GSE cases	658	54 (15-91)	71.9
English controls	1185	48	50
Dutch IBD cases[b]	588	46 (17-89)	57.7
Dutch Crohn's disease cases	298	44 (17-85)	70.5
Dutch ulcerative colitis cases	290	48 (19-89)	44.6

a Used in the comprehensive screen.
b Is the combination of the Crohn's disease and the ulcerative colitis group

Supplementary table 3. Tag SNP selection per gene

Gene name	Standard tagging	Alternative tagging	Gene name	Standard tagging	Alternative tagging
UMCU1	x		*UMCU22*		x
UMCU2	x		*UMCU23*		x
UMCU3	x		*UMCU24*	x	
UMCU4	x		*UMCU25*	x	
UMCU5	x		*UMCU26*	x	
UMCU6	x		*UMCU27*	x	
UMCU7	x		*UMCU28*		x
UMCU8	x		*UMCU29*		x
UMCU9	x		*UMCU30*	x	
UMCU10	x		*UMCU31*	x	
UMCU11	x		*UMCU32*	x	
UMCU12	x		*UMCU33*	x	
UMCU13	x		*UMCU34*	x	
UMCU14	x		*UMCU35*	x	
UMCU15	x		*UMCU36*	x	
UMCU16	x		*UMCU37*	x	
UMCU17	x		*UMCU38*	x	
UMCU18	x		*UMCU39*	x	
UMCU19	x		*UMCU40*	x	
UMCU20	x		*UMCU41*	x	
UMCU21		x			

Supplementary table 5. Chi-square comparison of the Set 1 controls versus the Set 2 controls. No stratification is observed

	Set 1 controls		Set 2 controls		Set 1 controls %		Set 2 controls %			
Rs number	Min allele	Maj allele	Min allele	Maj allele	Min allele	Maj allele	Min allele	Maj allele	Chi-square	P-value
rsUMCU172	402	510	380	540	44.1	55.9	43.0	57.0	0.218	0.6406
rsUMCU173	276	636	285	597	30.3	69.7	32.3	67.7	0.877	0.3492
rsUMCU117	418	494	422	466	45.8	54.2	47.5	52.5	0.516	0.4726
rsUMCU129	338	574	347	541	37.1	62.9	39.1	60.9	0.775	0.3787
rsUMCU140	362	550	377	513	39.7	60.3	42.4	57.6	1.324	0.2499

SECTION IV

CANDIDATE GENE AND GENE FAMILY STUDIES

Chapter 8

The interferon gamma gene in coeliac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility

MC Wapenaar, MJ van Belzen, JH Fransen, A Fariña Sarasqueta, RHJ Houwen, JWR Meijer, CJJ Mulder, C Wijmenga

J Autoimmunity 2004; 23:183-190

(reproduced with permission from Elsevier)

ELSEVIER

Journal of Autoimmunity 23 (2004) 183–190

journal of autoimmunity

www.elsevier.com/locate/issn/08968411

The interferon gamma gene in celiac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility

Martin C. Wapenaar[a,*,1], Martine J. van Belzen[a,b,1], Justin H. Fransen[a], Aranzazu Fariña Sarasqueta[a], Roderick H.J. Houwen[b], Jos W.R. Meijer[c], Chris J.J. Mulder[c,d], Cisca Wijmenga[a]

[a]*Department of Biomedical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands*
[b]*Pediatric Gastroenterology, University Medical Center Utrecht, Utrecht, The Netherlands*
[c]*Department of Pathology, Rijnstate Hospital, Arnhem, The Netherlands*
[d]*Department of Gastroenterology, VU Medical Center, Amsterdam, The Netherlands*

Received 21 January 2004; revised 12 May 2004; accepted 24 May 2004

Abstract

Celiac disease (CD) is a complex genetic disorder characterized by gluten intolerance. The Th1 immune response, with a key position for interferon gamma (IFN-γ), is an important determinant of intestinal remodeling in CD. We aimed at further ascertaining the role of IFN-γ, either as a genetic factor in the etiology, or as a facilitator of disease initiation/progression. Duodenal biopsies were sampled across distinct histopathological stages of the disease, including refractory CD (RCD), and used to determine IFN-γ gene (*IFNG*) expression by real-time RT-PCR. *INFG* expression correlated with the extent of tissue restructuring, reaching a 240-fold higher expression in total villous atrophy compared to healthy tissue. CD and RCD patients with similar lesions had comparable expression levels. Interestingly, patients in complete remission still had 7.6-fold residual over-expression. An *INFG* marker was tested in three cohorts of Dutch patients for both genetic linkage and association. Linkage analysis yielded no significant scores for *IFNG* or its flanking markers. In addition, *IFNG* allele frequencies were not differently distributed between cases and controls. Likewise, all alleles were randomly transmitted to affected children in parents-case trios. There is no evidence for *IFNG* as a predisposing gene in CD, despite its enhanced expression in patients in complete remission.

Keywords: Celiac disease; Gene expression; Genetics; Interferon gamma; Intestinal mucosa

1. Introduction

In celiac disease (CD; OMIM 212750), the ingestion of gluten provokes a combined Th1 and Th2 adaptive immune response [1]. The Th1 reaction is generally considered responsible for the long-term tissue remodeling observed in the small intestine [2]. The Th2 response is transient and results in the circulation of serum antibodies that are particularly useful in CD diagnosis and in monitoring compliance to a gluten-free diet [3,4]. The Th1 response is initiated by the binding of gluten peptides to the HLA-DQ2 and -DQ8 molecules on the surface of antigen presenting cells and subsequent recognition by CD4+ cells in the lamina propria [5]. Details on how these gluten peptides cross the epithelial boundary, however, remain obscure. Epitope binding to the DQ2 and DQ8 molecules is further enhanced by deamidation of gliadin by the enzyme tissue-transglutaminase, yielding an increased immune response [6,7].

* Corresponding author. Department of Biomedical Genetics, Stratenum 2.106, University Medical Center Utrecht, P.O. Box 80030, 3508 TA Utrecht, The Netherlands. Tel.: +31-30-253-8790; fax: +31-30-253-8479.
E-mail address: m.c.wapenaar@med.uu.nl (M.C. Wapenaar).
[1] Both authors contributed equally.

0896-8411/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jaut.2004.05.004

The cytokine interferon gamma (IFN-γ) plays a pivotal role in the subsequent T cell activation and concomitant tissue restructuring. In patients, the level of IFN-γ gene (*IFNG*) expression as well as the number of IFN-γ producing cells are increased, as observed for both the intestinal mucosa [8–11], and peripheral blood lymphocytes [10,12]. The in vitro gluten challenge of biopsies from CD patients in complete remission on a gluten-free diet (GFD) yielded *IFNG* mRNA levels comparable to that of untreated patients [9]. Organ cultures of duodenal biopsies from normal controls produced damaged enterocytes when exposed to purified INF-γ or culture supernatants from gluten-activated T cell clones. This toxic effect of the supernatants could be blocked by pre-incubation with anti-INF-γ antibodies [13]. In active CD there is a marked shift of the *INFG* expression from the lamina propria towards the epithelium, which may further impair the integrity of the epithelial boundary [11,14]. Moreover, villous epithelial damage in mice can be induced in vivo by injection of INF-γ [15]. Taking into account the crucial position of INF-γ in the Th1 pathway, it is plausible that certain polymorphisms of the *INFG* gene may lead to altered expression levels, and thus confer susceptibility for developing CD. Indeed, it was reported that an elevated level of in vitro IFN-γ production was associated with a specific allelic variant of the *IFNG* gene [16]. Celiac disease is a complex genetic disorder in which the major contributing genes—*HLA-DQ* [17] and a recently identified locus on 19p13 in the Dutch population [18]—attribute an estimated 60% of the total genetic risk. This implies that there must be other predisposing genes, although each of them will have only a limited effect. Considering its involvement in celiac disease, *IFNG* might be among those candidate genes that have not been identified yet.

In this report we followed *IFNG* expression in duodenal biopsies during successive stages of tissue remodeling in CD. To the best of our knowledge, this is the first study that correlates the level of *IFNG* expression with the extent of mucosal damage. We also tested our hypothesis that *IFNG* is a candidate gene by performing linkage and association studies in Dutch celiac disease patients.

2. Material and methods

2.1. Patient material for expression analysis

Patients visited the Rijnstate Hospital, Arnhem, which is a recognized center of expertise on celiac disease. They came either for a first-time examination ($n = 7$), or for a follow-up to monitor their response to a gluten-free diet ($n = 21$). Over a period of two years, 29 duodenal biopsies were sampled from 28 patients using forceps-spike endoscopy (Table 1). All biopsies were examined by the same pathologist (JWRM) and classified according to Marsh [2], using the UEGW criteria [19]. In parallel to the biopsies that were taken for histological analysis, two additional biopsies were sampled during the same session and snap-frozen in liquid nitrogen for expression analysis. The pathology report that was based on multiple biopsy samples was also considered representative for the biopsies used for the expression assays. Most patients were adults and they had an average age of 47 years. Five females (average age 48 years) with normal histology (M0), who had no history of CD, were used as controls. Seven patients were referred as new cases and had not started treatment. The 21 patients, who had been previously diagnosed with CD and initially presented with a MIII lesion, were all on a gluten-free diet (GFD). All patients who were on a GFD were negative for EmA antibodies, while patients not yet treated were EmA positive. The follow-up biopsies of the CD patients showed various levels of improvement from MI–MII up to complete remission (M0). A group of ten patients failed to recover despite strict adherence to a GFD. After exclusion of all factors that may cause a celiac-like phenotype, these patients were diagnosed as having refractory celiac disease (RCD) [20]. This latter group was subdivided into those with normal T cells (type I) and those with aberrant T cells (type II) [19,20]. All patients who volunteered for this study signed an informed consent. The study was approved by the Medical Ethics Committee (METC) of the University Medical Center Utrecht.

2.2. RNA isolation

Total RNA was isolated from two pooled snap-frozen biopsies (10–15 mg) in 500 μl TRIzol using the manufacturer's instructions (Gibco/Life Technologies, Rockville, MD). The biopsies were homogenized in the presence of TRIzol with 1 mm diameter glass beads by using a Mini-BeadBeater (BioSpec Products, Inc, Bartlesville, OK). We assessed the integrity and quantity of the RNA samples with Lab-on-a-Chip technology on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). Biopsies yielded on average 3.6 μg of total RNA/mg of tissue.

2.3. Real-time reverse transcription PCR

Relative quantification of the *IFNG* gene transcription in biopsies was determined by real-time RT-PCR. All reagents and analytical equipment used for this assay were purchased from Applied Biosystems (Foster City, CA) and used according to their recommendations. cDNA was synthesized from 1 μg total RNA in a 50 μl volume using the High Capacity Archive Kit on

Table 1
Details of patient and biopsy samples used in the expression study

ID	Diagnosis	Marsh class[a]	Gender	Age	GFD	HLA[b]
102	Normal	M0	F	30	–	DQ2/DQ8
104	Normal	M0	F	45	–	X/X
200	Normal	M0	F	34	–	DQ2/X
208	Normal	M0	F	70	–	NA
239	Normal	M0	F	59	–	NA
105	CD	M0	F	64	+	DQ2/DQ2
178	CD	M0	M	24	+	DQ2/DQ2
190	CD	M0	F	28	+	DQ2/DQ8
216	CD	M0	F	42	+	DQ2/X
218	CD	M0	F	72	+	NA
88	CD	MI	M	50	+	DQ2/DQ8
101	CD	MI	F	41	+	DQ2/X
197	CD	MI–MII	F	64	+	DQ2/X
205	CD	MI–MII	F	76	+	DQ2/X
54	CD	MII	F	43	+	DQ2/X
109	CD	MII	F	34	+	DQ2/DQ8
184[c]	CD	MII	F	19	–	DQ2/DQ2
194[d]	CD	MII	M	52	+	DQ2/X
85[d]	CD	MIIIa	M	51	+	DQ2/X
221	CD[e]	MIIIa	M	37	–	DQ2/X
100	CD[e]	MIIIb	F	15	–	DQ2/X
180	CD[e]	MIIIb	M	2	–	DQ2/X
259	CD[e]	MIIIc	F	54	–	DQ2/X
284	CD[e]	MIIIc	F	1	–	DQ2/X
215	RCD type I	MII–MIIIa	F	67	+	DQ2/DQ8
98	RCD type I	MIIIa	F	49	+	DQ2/X
214	RCD type I	MIIIa	F	83	+	DQ2/X
230	RCD type I	MIIIa	F	60	+	DQ2/X
228	RCD type I	MIIIb–c	F	27	+	DQ2/DQ8
86	RCD type II	MIIIa	F	66	+	DQ2/DQ2
162	RCD type II	MIIIa	F	50	+	DQ2/X
177	RCD type II	MIIIa	F	71	+	DQ2/X
288	RCD type II[e]	MIIIa	M	62	–	DQ2/X
199	RCD type II	MIIIb	M	62	+	DQ2/X

[a] Marsh classification: M0 (normal); MI (lymphocytosis); MII (additional crypt hyperplasia); MIIIa–c (additional villous atrophy; either partial [a], subtotal [b], or total [3]).
[b] X indicates other type than DQ2 or DQ8; NA: not available.
[c] Case 184 was not on GFD and was diagnosed CD MII through early intervention and improved after starting on a GFD.
[d] Samples 85 and 194 were taken from the same patient on a GFD over a one-year interval.
[e] Presented as new cases.

a GeneAmp PCR System 9700. *IFNG* transcripts were detected with Assay-on-Demand Hs00174143m1, which consists of a fluorescent Taqman probe specific for exon 2 and two flanking, exon–intron boundary-spanning primers (sequences are proprietary information held by the supplier). PCR cycling was performed on a 7900HT Sequence Detection System using 30 ng of reverse transcribed total RNA in 25 µl Universal PCR Master Mix. All samples were tested in triplicate. Transcripts of the housekeeping gene *GUSB* (detected by PDAR 4326320E) were used as an endogenous reference to correct for unwanted variability in total cDNA input between samples. Real-time amplification plots from the *IFNG* and *GUSB* transcripts were then used to extract Ct-values using SDS2.0 software. The $2^{-\Delta\Delta Ct}$ method [21] was applied for the conversion to relative expression levels by using the average ΔCt-value from the five normal controls as a calibrator.

2.4. Patients in this genetic study

Dutch Caucasian CD patients were collected throughout the Netherlands. The linkage study was performed on 84 affected sibpairs derived from 67 families. Details on the families in this cohort have been described elsewhere [18]. Genetic studies of the *IFNG* gene were performed on three cohorts. The first cohort consisted of 207 independent CD cases and 210 independent controls. The cases and controls both had a mean age of 39 years and a male-to-female ratio of 35:65. The second cohort comprised 122 parents-case trios and the patients in this cohort had a mean age of 17 years and also a gender ratio of 35:65. The third cohort was the same affected sibpair cohort as used in the linkage analysis. To ensure that all patients met the same strict clinical criteria, the histology on their biopsies was re-examined by the same pathologist

(JWRM). Only patients and affected siblings who met the MIII criteria [2] were included in our study. Blood samples were collected and DNA was isolated according to standard laboratory procedures [18]. A written informed consent was obtained from all participating patients and family members. The METC of the University Medical Center Utrecht approved this study.

2.5. The IFNG marker with flanking region

PCR fragments containing the CA-repeat polymorphism in the first intron of the *IFNG* gene were obtained in a reaction volume of 10 µl containing 25 ng of DNA, 0.2 mM dNTPs, 2.5 mM $MgCl_2$, 50 ng of forward primer 5′-TTA TTC TTA CAA CAC AAA ATC AAA TC-3′, 50 ng of reverse, fluorescence-labeled primer 5′-ATA CAA AAA CAA AAA ACA GCA AAG C-3′ and 0.4 U AmpliTaq Gold. The PCR products of 190−200 bp were separated on a 3700 DNA sequencer and analyzed by Genescan 3.5 and Genotyper 2.0 software. Reagents, sequence detector, and software were from Applied Biosystems (Foster City, CA). The nomenclature of the alleles is adopted from Pravica et al. [16]; alleles 1−5 correspond to 11−15 CA-repeats, respectively. In our sample, allele 1 was not present, and alleles 6 and 7 (16 and 17 CA-repeats, respectively) had frequencies below 0.5% and were therefore not included in the analysis. In addition, the following five *IFNG* flanking markers on 12q15 and spanning ~27 cM were analyzed in the genetic study: D12S398, D12S1294, D12S375, D12S1052, D12S1064. All genotypes were checked independently by two researchers.

2.6. Genetic linkage analysis of the IFNG gene region

The *IFNG* CA-repeat and the five flanking markers were tested for linkage to celiac disease by non-parametric linkage analysis. The analysis was performed using the MAPMAKER/SIBS program [22], both single and multipoint analyses were carried out. Marker positions and order were based on the Marshfield and Ensembl genetic maps (http://research.marshfieldclinic.org/genetics/; http://www.ensembl.org). Allele frequencies were calculated from the entire dataset as described previously [18].

2.7. Genetic association studies of the IFNG gene

Association analysis of the *IFNG* gene was performed on three cohorts. In the first cohort allele and genotype frequencies were compared between cases and controls and *P*-values were obtained by χ^2 analysis. The second (trios) and third cohorts (sibpairs) were subjected to transmission/disequilibrium tests (TDT). Alleles transmitted and not-transmitted to affected offspring in the parent-case trios were compared using a χ^2 test with 1 degree of freedom. In addition, the two markers flanking *IFNG*, D12S1294 and D12S375, were also subjected to TDT analysis in the second and third cohorts.

3. Results

3.1. IFNG transcription profiling

Information on the patients and controls used in this *IFNG* expression study is summarized in Table 1. The results of the *INFG*-transcript quantification by real-time RT-PCR on the duodenal biopsies are presented in Fig. 1. The relative expression in all individual biopsies was obtained after normalization against the average value of the five normal controls. Also indicated is the trend line connecting the average expression of each patient's category. All the patients' categories, with the exception of the CD M0 group ($P = 0.07$), had average expression levels that were significantly elevated compared to the average value for the normal controls ($P < 0.05$). Two features of the *IFNG* expression profiling stand out. First, there is considerable expression variability among samples within the same category, and second, superimposed on this variability there is a remarkable difference between the averages of each category. This observed increase of the average *IFNG* transcription correlates well with the extent of tissue remodeling in the duodenum. The variability among the samples was not confined to just the patients but was also observed in the normal control group. These control biopsies varied in *IFNG* transcription from 0.3-fold (case 208) to 2.8-fold (case 200) the average value in this class, indicating a maximum 9.3-fold difference in expression between these two extremes. A similar pattern was observed in the patient samples, be it with some more pronounced differences that could be attributed to just a few patients (e.g. cases 109 and 216). In general, individual biopsy samples varied in *IFNG* transcription within a range of two- to three-fold above or below the average expression value of their category. Despite this inter-individual variability, the average *IFNG* expression of the various patient categories appeared to reflect the extent of the tissue lesion. Partial villous atrophy (CD MIIIa) shows an average 30-fold over-expression that increases to 60-fold in sub-total (CD MIIIb), and to a 240-fold up-regulation in the transition to complete atrophy of the villi (CD MIIIc). In the CD MII and MI classes this value fluctuates between 35- and 40-fold, similar to that in CD MIIIa. Interestingly, although the CD-remission patients had a histology indistinguishable from the normal controls they still produced 7.6-fold more *IFNG* transcripts on

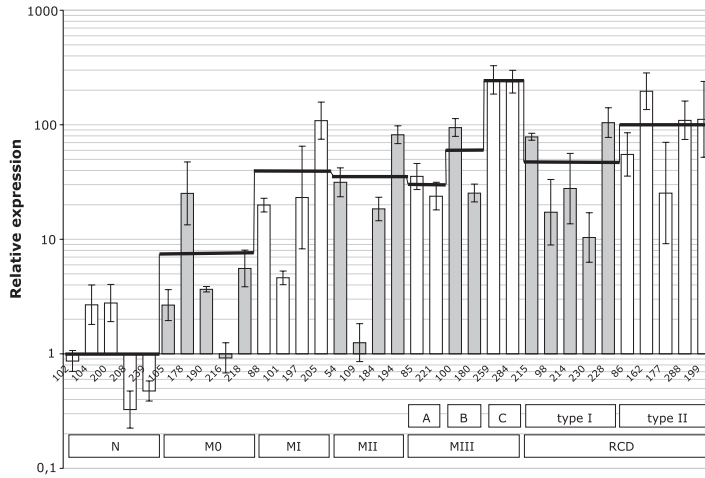


Fig. 1. Relative quantification of *IFNG* transcription at successive stages of tissue remodeling in duodenal biopsies from CD and RCD patients. Real-time RT-PCR data from *IFNG* and the endogenous reference gene *GUSB* were processed using the $2^{-\Delta\Delta Ct}$ method [21]. The relative expression for each individual biopsy is normalized to the averaged value of the five normal controls (N). Numbers directly below the bars refer to the biopsies as summarized in Table 1. Biopsies that belong to the same class, as indicated in the lower part of the graph, have similar fill patterns in the histogram. A solid black line connects the average values from all categories to illustrate the correlation between *IFNG* expression and the extent of tissue remodeling. Note that the relative expression is displayed on a logarithmic scale. Significant differences in expression were tested for using a *T*-test (two-sided, assuming equal variance), either with or without a Bonferroni correction for multiple testing. Without the conservative correction, the differences between the following classes were significant ($P < 0.05$): N vs. MI, MII, MIIIa/b/c, RCD I/II; M0 vs. MIIIc; MIIIa vs. MIIIc; and MIIIc vs. RCD I. With the correction applied the significant differences ($P < 0.05$) were limited to N vs. MIIIc, RCD I, RCD II. The expression ranges observed for the various patient categories were: CD M0 (0.9–25); CD MI (4.6–109); CD MII (1.3–82); CD MIIIa (24–35); CD MIIIb (25–95); CD MIIIc (238–246); RCD type I (10–104); and RCD type II (25–195). More pronounced differences resulted from a few outliers: cases 109 (25-fold below CD MII average), 216 (8.4-fold below CD M0 average), 230 (4.8-fold below RCD type I average), and 177 (4.0-fold below RCD type II average).

average. This could suggest that these patients had either not completely recovered, or have a genetically determined enhanced *IFNG* expression. Finally, the patients refractory to a gluten-free diet (GFD), both in RCD type I (average 50-fold) and type II (average 100-fold), showed no pronounced difference in *IFNG* expression compared to CD patients with a similar histology.

3.2. Genetic linkage analysis

Genetic linkage between the *IFNG* microsatellite marker and CD was tested in 84 affected sibpairs from 67 independent families. In addition, five flanking marker loci were included spanning a total region of 26.8 cM (37.6 Mb) on chromosomal band 12q24.1. The results from both the single-point (Table 2) and multi-point analyses (data not shown, and Ref. [18]) indicated no evidence for genetic linkage between the *IFNG* region and CD.

3.3. Genetic association analyses

The overall distribution of *IFNG* CA-repeat alleles was not significantly different between cases and controls (Table 3). The influence of allele 2 (12 CA-repeats) was specifically investigated, as this allele was reported to be associated with high in vitro expression of INFG [16]. Homozygosity or heterozygosity for this allele did not significantly contribute to disease risk

Table 2
Maximum likelihood scores (MLS) for linkage between CD and the *INFG* region

Locus	Genetic map location (cM)	Physical map location (Mb)	MLS
D12S398	68.2	53.7	0.0
D12S1294	78.0	67.7	0.0
IFNG	79.2	68.4	0.0
D12S375	80.5	68.7	0.0
D12S1052	83.2	75.5	0.0
D12S1064	95.0	91.3	0.8

(Table 3). Likewise, no significant distortion of random transmission of CA-repeat alleles to affected offspring was present in the parent-case trios (Table 4) or the affected sibpairs (data not shown). Therefore, none of the cohorts provided any evidence for genetic association between *IFNG* and CD. Additional TDT analysis of the *IFNG*-flanking markers, D12S1294 and D12S375, in the trios and sibpair cohorts also provided no evidence for association.

Table 4
Transmission of *IFNG* CA-repeat alleles to celiac disease patients

Allele	Transmitted[a]	Not transmitted
2	51	51
3	47	51
4	14	12
5	7	6

[a] None of the alleles were significantly overtransmitted to affected offspring.

4. Discussion

The increase of *IFNG* expression in untreated CD patients compared to normal controls and CD patients who recovered on a GFD has been well documented [8–11]. To the best of our knowledge, no accurate measurement of *IFNG* transcription has been reported on CD patients who were stratified according to the successive Marsh stages. The data we have presented shows the correlation between the average level of *IFNG* expression and the extent of tissue restructuring in the intestinal mucosa. In biopsies with complete atrophy of the villi (MIIIc), this expression is about 240-fold higher than that measured in the averaged controls.

Despite this clear phenotype-expression correlation, we also observed considerable expression variability within all Marsh classes, although most samples varied within a range of two- to three-fold the average value of their category. There were a few outliers but the cause for their extreme transcriptional behavior could not be derived from any of the patients' features as summarized in Table 1. In these cases it is also unlikely to be the result of lesion heterogeneity since their expression value did not fit within the range of the adjacent Marsh class. A recent study on jejunal T cells in CD showed a comparable variability in *IFNG* expression [11].

Table 3
Allele and genotype frequencies of the *IFNG* CA-repeat polymorphism in cases and controls

	CD cases (N = 207)	Controls (N = 210)
Allele[a]		
2	0.46	0.49
3	0.44	0.41
4	0.06	0.06
5	0.04	0.04
Genotype[a]		
2,2	0.23	0.22
2,3	0.37	0.41
2,4	0.04	0.08
2,5	0.04	0.04
2,7	0.004	0.0
X,X[b]	0.32	0.25

[a] No significant differences in either allele or genotype frequencies were present between cases and controls.
[b] Allele X is not allele 2.

However, it cannot be excluded that some of this expression variation resulted from the biopsy sampling. Two samples from one patient were collected over a one-year interval while he was on GFD (biopsies 85 and 194). Despite the modest improvement from MIIIa to MII, the *IFNG* expression increased 2.3-fold. Possibly, lesion heterogeneity in the duodenum and non-representative biopsy sampling could explain this skewed expression between the marginally different MII and MIIIa conditions. On the other hand, cytokines like IFN-γ are part of an interactive network and they influence each other's expression, as much as they are controlled by the intrinsic features of the genes themselves [16]. Therefore, the observed expression variation could result from polymorphisms in the *IFNG* gene, or in the genes that regulate its expression. Given this remarkable expression variability between individuals, it would be interesting to identify the genetic factors responsible by genotype-expression comparison and to determine if some of these genotypes pose a health risk.

In patients on a GFD and in complete remission (M0), *IFNG* expression dropped sharply but was on average still 7.6-fold higher than in the normal controls. A similar observation was made for childhood CD [11]. This residual activity might suggest that persistent traces of an inflammation are still present in an otherwise normalized mucosa. These patients might still be encountering trace amounts of contaminating gluten in their diet, despite a strict adherence to their GFD. Also sensitization to other natural antigens in the aftermath of the gluten-induced enteropathy cannot be excluded completely. However, the negative EmA antibody tests, and the total absence of any clinical features in these recovered patients do not support this option. Alternatively, these individuals may carry a genetic factor that increases steady-state *IFNG* expression and predisposes to develop an autoimmune disorder like CD. This, together with the observed expression variation, and the important position in the Th1 pathway, prompted us to investigate the *IFNG* gene as a causal factor in CD pathogenesis. We tested a microsatellite marker from within the *IFNG* gene, of which a certain allele was reported to be associated with high in vitro expression of IFN-γ [16]. This CA-repeat polymorphism was also

reported to be associated with several autoimmune disorders such as type 1 diabetes mellitus [23,24], rheumatoid arthritis [25], and multiple sclerosis [26,27]. However, neither our sibpair analysis nor the three association studies provided any evidence that *IFNG* is a predisposing gene for CD. It is unlikely that *IFNG* makes a strong contribution to CD in other populations since no evidence for linkage to the 12q15 region was found in any of the other genome-wide screens performed so far (reviewed by Van Belzen et al. [18]). Nevertheless, we cannot exclude the possibility that enhanced *IFNG* transcription in treated celiac disease patients results from a polymorphic gene that regulates the expression of *IFNG*. Possible candidates like *IL12B* and *IRF-1*, however, failed to show association with CD in the Dutch population [28].

When we compared *IFNG* expression between RCD and CD patients with similar lesions, we found no obvious differences. This suggests that the level of *IFNG* transcription reflects the extent of tissue remodeling regardless of any molecular changes that drive the transition to the refractory condition. However, the two-fold average increase associated with the switch from RCD type I to type II shown in Fig. 1 might be provoked by the presence of the cryptic T-cell lymphoma in the latter [29].

We have demonstrated that the extent of mucosal remodeling in CD is correlated with an enhanced level of *IFNG* transcription that shows considerable fluctuation between individuals. Despite its crucial role in the Th1 pathway, and the increased expression in recovered patients, there is no evidence that the *IFNG* gene itself confers susceptibility for developing CD. Based on the results from our genome-wide expression study, we recently proposed a model that explains the pathology of CD by an increased proliferation-without-differentiation [30]. In this model the elevated *IFNG* expression could induce or sustain the undifferentiated but highly proliferative condition of the celiac mucosa, rather than causing tissue damage by apoptosis and necrosis.

Acknowledgements

We wish to thank the celiac disease patients and their families for participating in this study. Results of the HLA-typing were kindly provided by Dr. B. Crusius (Laboratory of Immunogenetics, VU Medical Center, Amsterdam). Alfons Bardoel and Ineke Lavrijsen assisted in genotype ascertainment, and Lude Franke in the statistical analysis of the expression data (members of Department of Biomedical Genetics, UMC Utrecht). We thank Dr. F. Koning for critically reading this report and are indebted to Jackie Senior for editing the manuscript. This study was supported by grants WS 00-13 and WS 97-44 from the Dutch Digestive Diseases Foundation (MLDS).

References

[1] Schuppan D. Current concepts of celiac disease pathogenesis. Gastroenterology 2000;119:234–42.
[2] Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). Gastroenterology 1992;102:330–54.
[3] Fasano A, Catassi C. Current approaches to diagnosis and treatment of celiac disease: an evolving spectrum. Gastroenterology 2001;120:636–51.
[4] Green PH, Jabri B. Coeliac disease. Lancet 2003;362:383–91.
[5] Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. Nat Rev Immunol 2002;2:647–55.
[6] van de Wal Y, Kooy Y, van Veelen P, Pena S, Mearin L, Papadopoulos G, et al. Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. J Immunol 1998;161:1585–8.
[7] Molberg O, McAdam SN, Korner R, Quarsten H, Kristiansen C, Madsen L, et al. Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. Nat Med 1998;4:713–7.
[8] Kontakou M, Sturgess RP, Przemioslo RT, Limb GA, Nelufer JM, Ciclitira PJ. Detection of interferon gamma mRNA in the mucosa of patients with coeliac disease by in situ hybridisation. Gut 1994;35:1037–41.
[9] Nilsen EM, Jahnsen FL, Lundin KE, Johansen FE, Fausa O, Sollid LM, et al. Gluten induces an intestinal cytokine response strongly dominated by interferon gamma in patients with celiac disease. Gastroenterology 1998;115:551–63.
[10] Lahat N, Shapiro S, Karban A, Gerstein R, Kinarty A, Lerner A. Cytokine profile in coeliac disease. Scand J Immunol 1999;49:441–6.
[11] Forsberg G, Hernell O, Melgar S, Israelsson A, Hammarstrom S, Hammarstrom ML. Paradoxical coexpression of proinflammatory and down-regulatory cytokines in intestinal T cells in childhood celiac disease. Gastroenterology 2002;123:667–78.
[12] Hansson T, Dannaeus A, Klareskog L. Cytokine-producing cells in peripheral blood of children with coeliac disease secrete cytokines with a type 1 profile. Clin Exp Immunol 1999;116:246–50.
[13] Przemioslo RT, Lundin KE, Sollid LM, Nelufer J, Ciclitira PJ. Histological changes in small bowel mucosa induced by gliadin sensitive T lymphocytes can be blocked by anti-interferon gamma antibody. Gut 1995;36:874–9.
[14] Olaussen RW, Johansen FE, Lundin KE, Jahnsen J, Brandtzaeg P, Farstad IN. Interferon-gamma-secreting T cells localize to the epithelium in coeliac disease. Scand J Immunol 2002;56:652–64.
[15] Guy-Grand D, DiSanto JP, Henchoz P, Malassis-Seris M, Vassalli P. Small bowel enteropathy: role of intraepithelial lymphocytes and of cytokines (IL-12, IFN-gamma, TNF) in the induction of epithelial cell death and renewal. Eur J Immunol 1998;28:730–44.
[16] Pravica V, Asderakis A, Perrey C, Hajeer A, Sinnott PJ, Hutchinson IV. In vitro production of IFN-gamma correlates with CA repeat polymorphism in the human IFN-gamma gene. Eur J Immunogenet 1999;26:1–3.
[17] Pena AS, Wijmenga C. Genetic factors underlying gluten-sensitive enteropathy. Curr Allergy Asthma Rep 2001;1:526–33.
[18] Van Belzen MJ, Meijer JW, Sandkuijl LA, Bardoel AF, Mulder CJ, Pearson PL, et al. A major non-HLA locus in celiac disease maps to chromosome 19. Gastroenterology 2003;125:1032–41.

[19] Report of a working group of the United European Gastroenterology Week in Amsterdam. When is a coeliac a coeliac? Eur J Gastroenterol Hepatol 2001;13:1123–8.
[20] Goerres MS, Meijer JW, Wahab PJ, Kerckhaert JA, Groenen PJ, Van Krieken JH, et al. Azathioprine and prednisone combination therapy in refractory coeliac disease. Aliment Pharmacol Ther 2003;18:487–94.
[21] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 2001;25:402–8.
[22] Kruglyak L, Lander ES. Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet 1995;57: 439–54.
[23] Awata T, Matsumoto C, Urakami T, Hagura R, Amemiya S, Kanazawa Y. Association of polymorphism in the interferon gamma gene with IDDM. Diabetologia 1994;37:1159–62.
[24] Jahromi M, Millward A, Demaine A. A CA repeat polymorphism of the IFN-gamma gene is associated with susceptibility to type 1 diabetes. J Interferon Cytokine Res 2000;20:187–90.
[25] Khani-Hanjani A, Lacaille D, Hoar D, Chalmers A, Horsman D, Anderson M, et al. Association between dinucleotide repeat in non-coding region of interferon-gamma gene and susceptibility to, and severity of rheumatoid arthritis. Lancet 2000;356: 820–5.
[26] Vandenbroeck K, Opdenakker G, Goris A, Murru R, Billiau A, Marrosu MG. Interferon-gamma gene polymorphism-associated risk for multiple sclerosis in Sardinia. Ann Neurol 1998;44:841–2.
[27] Goris A, Epplen C, Fiten P, Andersson M, Murru R, Sciacca FL, et al. Analysis of an IFN-gamma gene (IFNG) polymorphism in multiple sclerosis in Europe: effect of population structure on association with disease. J Interferon Cytokine Res 1999;19: 1037–46.
[28] Seegers D, Borm ME, van Belzen MJ, Mulder CJ, Bailing J, Crusius JB, et al. IL12B and IRF1 gene polymorphisms and susceptibility to celiac disease. Eur J Immunogenet 2003;30: 421–5.
[29] Ryan BM, Kelleher D. Refractory celiac disease. Gastroenterology 2000;119:243–51.
[30] Diosdado B, Wapenaar MC, Franke L, Duran KJ, Goerres MJ, Hadithi M, et al. A microarray screen for novel candidate genes in coeliac disease pathogenesis. Gut 2004;53:944–51.

Chapter 9

IFNG association with coeliac disease in the Dutch population

MC Wapenaar, A Zhernakova, AJ Monsuur,
ML Mearin, CJJ Mulder, and C Wijmenga

(Submitted for publication)

ABSTRACT

The intestinal inflammation in coeliac disease (CD) is characterized by an immune response in which IFN-γ plays a pivotal role. The CA-repeat in intron 1 of *IFNG* has been genetically associated with autoimmune diseases and is correlated with elevated *in vitro* IFN-γ production. In a previous study we were not able to find evidence for genetic association of this CA-repeat polymorphism in Dutch CD patients. Here we present five haplotype tagging SNPs that are in complete linkage disequilibrium and detect six common haplotypes in a region of 83 kb that includes *IFNG*. We observed significant association for the intron 3 variant rs2069718*A (P = 0.002; *OR* 1.32; 95% *CI* 1.11 – 1.57), in a Dutch cohort of 464 cases and 659 controls. We observed no association with the high-risk $(CA)_{12}$-allele, which also did not occur on rs2069718*A-containing haplotypes. Therefore, the variant rs2069718*A may represent an improved susceptibility marker for autoimmune diseases.

INTRODUCTION

Coeliac disease (CD) is a gluten-induced inflammatory condition of the small gut characterized by atrophy of the villi, crypt hyperplasia, and lymphocyte infiltration[1]. The disease is a complex genetic trait and the major susceptibility locus is comprised of the *HLA-DQA* and *–DQB* genes[2]. The encoded HLA-DQ2 and -DQ8 proteins on the surface of antigen presenting cells bind (deamidated) gluten-peptides and stimulate T cells in the lamina propria to display a Th1 phenotype[3]. The Th1 proinflammatory cytokine interferon-gamma (IFN-γ) plays a prominent role in CD pathogenesis. CD patients respond to gluten with an elevated IFN-γ level and an increase in the number of IFN-γ-producing cells in the gut mucosa[4, 5]. The potency of IFN-γ to remodel the mucosal architecture was demonstrated in *ex vivo* tissue explants[6], and in an *in vivo* rat model[7]. We reported earlier that *IFNG* mRNA levels correlated with progression of tissue restructuring in patients following a gluten-free diet[8], and that a vast number of genes up- and downstream in the *IFNG*-pathway behaved similarly (Diosdado, submitted). We also observed considerable *IFNG* expression variability between patients that might be caused by cis- or trans-acting genetic factors[8]. By regulating *IFNG* expression these very same factors could contribute to the genetic susceptibility for CD. The $(CA)_{12}$-allele of the microsatellite marker in the first intron of *IFNG* was reportedly associated with enhanced in vitro IFN-γ production[9]. Genetic association of this CA-repeat, and the adjacent +874 T/A variant which is in linkage disequilibrium (LD) with it, has been observed in several autoimmune disorders[10-14], including a study in Spanish CD patients[15]. However, in Dutch CD patients we could find no evidence for this association[8]. In the present study we used the LD structure of *IFNG* to select five haplotype tagging (ht) SNPs across the gene and we report significant genetic association for one of

these in our Dutch CD patient cohort. This htSNP may also prove to be of value for studying other autoimmune disorders.

RESULTS AND DISCUSSION

The LD structure of *IFNG*, based on HapMap (Phase II), is depicted in figure 1. The entire *IFNG* gene of 5.0 kb is contained within a region of complete LD (D' = 1). This haploblock spans more than 83 kb, stretches from the SNPs rs10467156 to rs12320174, and includes 171 SNPs which form five frequent haplotypes (> 5%). Five htSNPs were selected from this LD-block (figure 1) and tested for genetic association in a case/control cohort of Dutch CD patients[16] (see table 1). Two SNPs showed significant association: rs2069718*A-allele (*P* = 0.002, *OR* 1.32 95% *CI* 1.11 - 1.57) and rs2193049*C-allele (*P* = 0.017, *OR* 1.27 95% *CI* 1.04 - 1.54), while the rs10878760*T-allele showed a trend towards association (*P* = 0.06, *OR* 1.25 95% *CI* 0.99 - 1.59). When a Bonferroni correction was applied to correct for the number of tests (n=5), only the association with SNP rs2069718 remained significant.

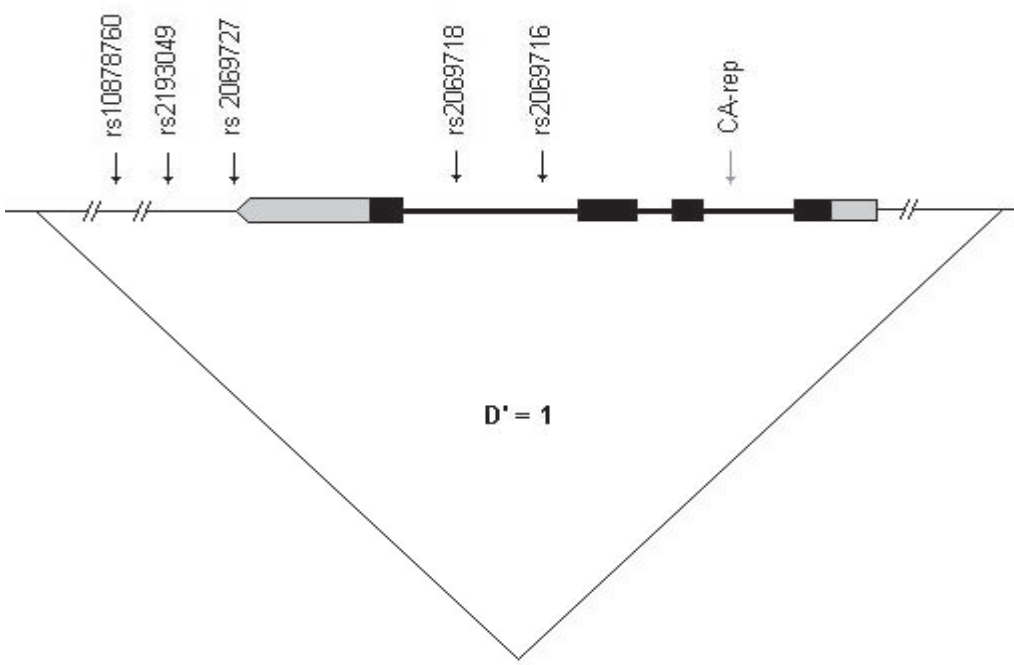


Figure 1. Genomic organization of the *IFNG* gene, the LD-block in which it is contained, and the five htSNPs and intron 1 CA-repeat that were use in this study. Solid boxes indicate exons with coding (black) and untranslated regions (grey) indicated. The lower portion of the figure shows the pairwise linkage-disequilibrium structure between indicated SNPs given by D' statistics based on the European population in the HapMap database (Phase II).

Table 1. Allele distribution of htSNPs from the *IFNG* locus in Dutch CD cases and controls

		Cases (n=464)		Controls (n=659)				
htSNP	Allele	Counts	Freq (%)	Counts	Freq (%)	P-value	OR	95% CI
rs10878760	G	775	84	1141	87	ref	1	
	T	149	16	175	13	0.060	1.25	0.99 - 1.59
rs2193049	G	608	71	955	76	ref	1	
	C	248	29	307	24	0.017	1.27	1.04 - 1.54
rs2069727	G	434	47	660	51	ref	1	
	A	488	53	644	49	0.100	1.15	0.97 - 1.36
rs2069718	G	530	59	847	66	ref	1	
	A	366	41	443	34	0.002	1.32	1.11 - 1.57
rs2069716	A	803	95	1190	94	ref	1	
	G	47	5	70	6	0.980	0.995	0.68 - 1.46

Genotypes were established on two different platforms: the Illumina BeadStation (rs2193049 and rs2069716) and the Applied Biosystems 7900HT (rs10878760, rs2069727, and rs2069718). The analysis on the Illumina platform was part of a comprehensive screen as described in detail by Monsuur, 2005. On the Applied Biosystems platform we used Taqman Assay-on-Demand (rs2069718) and Assay-by-Design (rs10878760, rs2069727). Protocols and reagents were obtained from Illumina and Applied Biosystems, respectively. Genotypes of all five SNPs were made from a total of 464 cases and 659 controls. Allele distribution in cases and controls were compared using the COCAPHASE module of the UNPHASED statistical package (Dudbridge, 2003). All htSNPs were in Hardy-Weinberg equilibrium in the cases and controls and calculated with the Haploview application (Barrett, 2005). Odds ratios were calculated and confidence intervals were approximated using Woolf's method with Haldane's correction (Haldane, 1956). Details on the htSNPs can be found in supplementary table 1.

We also re-examined the CA-repeat from intron 1 in our Dutch patient cohort, which has been expanded with 257 cases and 449 controls since our previous study[8] (table 2). The distribution of alleles was equal between cases and controls, and nearly identical to that reported earlier[8]. This also clearly demonstrated that population homogeneity was maintained after a major increase in our sample size. Again, the $(CA)_{12}$-allele that had been correlated with higher *in vitro* IFN-γ production was not associated with CD[9].

The five htSNPs identified six common *IFNG* haplotypes with frequencies > 5% (see table 3), but the overall haplotype distribution was not significantly different between cases and controls ($P = 0.17$). Only one of the three haplotypes that carried the associated rs2069718*A-allele (G-C-A-A-A, haplotype 4) was borderline significant ($P = 0.05$), although all three (haplotypes 3, 4, and 5) showed a similar trend of increased frequency in CD cases. As might be expected, we observed that the CA-repeat was in strong LD with the five htSNPs. Three out of the four frequent CA-alleles, $(CA)_{12}$, $(CA)_{14}$, and $(CA)_{15}$, were predominantly present on the most frequent G-G-G-G-A htSNP haplotype 1. The $(CA)_{13}$-allele, on the contrary, was distributed more or less equally among the haplotypes 2, 3, and 4

(table 4). Extending the htSNP haplotypes with the CA-repeat did not change the overall haplotype distribution among the cases and controls (P = 0.13).

Table 2. Allele distribution of the *IFNG* intron 1-CA-repeat in Dutch CD cases and controls

CA-repeat	Cases (n=464)		Controls (n=659)				
Allele	Counts	Freq (%)	Counts	Freq (%)	P-value	OR	95% CI
$(CA)_{12}$	421	46	639	49	ref	1	
$(CA)_{13}$	402	44	546	42	0.220	1.12	0.94 - 1.34
$(CA)_{14}$	64	7	67	5	0.040	1.45	1.01 - 2.08
$(CA)_{15}$	33	4	46	4	0.710	1.10	0.69 - 1.74

Generation of *IFNG* CA-repeat fragments by PCR, sizing on an ABI 3730 DNA sequencer, and genotype analysis using Genescan 3.5 and Genotyper 2.0 software was performed as described previously (Wapenaar, 2004). The $(CA)_{11}$-allele was not present in our cohorts, while the frequencies of the $(CA)_{16}$-, $(CA)_{17}$-, and $(CA)_{18}$-alleles were below 0.5% and were not included in the analysis. Allele distributions were calculated as indicated in table 1.

Table 3. Haplotype distribution of *IFNG* htSNPs in Dutch CD patients and controls

		Cases (n=464)		Controls (n=659)				
Haplotype	Number	Counts	Freq (%)	Counts	Freq (%)	P-value	OR	95% CI
G-G-G-G-A	1	339	42	544	45	ref		
G-G-A-G-A	2	104	13	186	15	0.70	0.90	0.68 - 1.18
T-C-A-A-A	3	130	16	168	14	0.40	1.24	0.95 - 1.62
G-C-A-A-A	4	109	13	132	11	0.05	1.33	1.00 - 1.77
G-G-A-A-A	5	86	11	123	10	0.75	1.12	0.83 - 1.53
G-G-G-G-G	6	45	6	62	5	0.80	1.17	0.78 - 1.75

The order of the htSNPS in the haplotype is: rs10878760 - rs2193049 - rs2069727 - rs2069718 - rs2069716. The overall *P*-value for the haplotypes is 0.170. Haplotype association was estimated using the UNPHASED package (Dudbridge, 2003).

The report that the $(CA)_{12}$-allele was associated with higher *in vitro* IFN-γ production has since its appearance[9], sparked a large number of genetic studies on its possible involvement in autoimmune diseases. Indeed, genetic associations have been reported for type 1 diabetes[10], rheumatoid arthritis[11], multiple sclerosis[12], and systemic lupus erythematosus[13]. However, the correlation between this microsatellite allele and IFN-γ production was not undisputed[17, 18]. In our previous study using both a case/control cohort and parents-affected offspring trios[8], and again confirmed here with the expanded case/control cohort, we could not find evidence for genetic association between this CA-repeat and CD. In a Spanish CD study association was observed for the $(CA)_{12}$-allele in familial transmission, but

Table 4. Combined *IFNG* htSNPs/CA-repeat haplotype distribution in Dutch CD patients and controls

	Cases (n=464)		Controls (n=659)				
Haplotype	Counts	Freq (%)	Counts	Freq (%)	P-value	OR	95% CI
G-G-G-G-A-$(CA)_{12}$	323	40	519	43	ref	1	
G-G-A-G-A-$(CA)_{13}$	101	12	182	15	0.42	0.89	0.68 – 1.18
T-C-A-A-A-$(CA)_{13}$	124	15	167	14	0.20	1.19	0.91 – 1.56
G-C-A-A-A-$(CA)_{13}$	105	13	130	11	0.08	1.30	0.97 – 1.74
G-G-G-G-G-$(CA)_{12}$	44	5.4	59	4.9	0.39	1.20	0.80 – 1.82
G-G-A-A-A-$(CA)_{14}$	52	6.4	59	4.9	0.09	1.41	0.95 – 2.11
G-G-A-A-A-$(CA)_{15}$	26	3.4	40	3.2	0.87	1.14	0.64 – 1.76

See legend of table 3 for details.

could not be replicated in a case/control cohort[15]. In the present study we have demonstrated association in the Dutch population between CD and the htSNP rs2069718*A-allele by using a haplotype tagging SNP approach. This SNP allele is present on three common haplotypes (numbers 3, 4, and 5) that do not carry the $(CA)_{12}$-allele. The disparity in the studies with the CA-repeat may be explained by the relative instability of microsatellite alleles, such that its alleles may occur on diverse haplotypic backgrounds in the various ethnic groups. It is unlikely that the associated SNP rs2069718 is a causative variant and it may turn out to be a daunting task to identify the true mutant given the strong LD within this locus. Evolutionary conserved noncoding sequence elements that regulate IFN-γ have been identified[19]. Unfortunately these regions are characterized by a lack of SNP variants. Identification of the true genetic variants that contribute to differential *IFNG* expression, and thus to autoimmune susceptibility, may require large-scale whole-gene sequencing. In conclusion, we demonstrated that 1) *IFNG* is associated with CD in the Dutch population, although it is a minor contribution to the overall disease susceptibility, and 2) the use of SNP variants, particularly rs2069718, is preferable over that of the first-intron CA-repeat.

ACKNOWLEDGEMENTS

We thank the Dutch coeliac disease patients and their family for participating in this study. We also thank Joachim Schweizer (LUMC) for assistance in collecting samples. This study was financially supported by the Celiac Disease Consortium, an Innovative Cluster approved by the Netherlands Genomics Initiative and partially funded by the Dutch Government (BSIK03009).

REFERENCES

1. Green PH, Jabri B. Coeliac disease. Lancet 2003;362:383-91.
2. van Heel DA, Hunt K, Greco L, et al. Genetics in coeliac disease. Best Pract Res Clin Gastroenterol 2005;19:323-39.
3. Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. Nat Rev Immunol 2002;2:647-55.
4. Kontakou M, Sturgess RP, Przemioslo RT, et al. Detection of interferon gamma mRNA in the mucosa of patients with coeliac disease by in situ hybridisation. Gut 1994;35:1037-41.
5. Forsberg G, Hernell O, Melgar S, et al. Paradoxical coexpression of proinflammatory and down-regulatory cytokines in intestinal T cells in childhood coeliac disease. Gastroenterology 2002;123:667-78.
6. Przemioslo RT, Lundin KE, Sollid LM, et al. Histological changes in small bowel mucosa induced by gliadin sensitive T lymphocytes can be blocked by anti-interferon gamma antibody. Gut 1995;36:874-9.
7. Stepankova R, Kofronova O, Tuckova L, et al. Experimentally induced gluten enteropathy and protective effect of epidermal growth factor in artificially fed neonatal rats. J Pediatr Gastroenterol Nutr 2003;36:96-104.
8. Wapenaar MC, van Belzen MJ, Fransen JH, et al. The interferon gamma gene in coeliac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility. J Autoimmun 2004;23:183-90.
9. Pravica V, Asderakis A, Perrey C, et al. In vitro production of IFN-gamma correlates with CA repeat polymorphism in the human IFN-gamma gene. Eur J Immunogenet 1999;26:1-3.
10. Jahromi M, Millward A, Demaine A. A CA repeat polymorphism of the IFN-gamma gene is associated with susceptibility to type 1 diabetes. J Interferon Cytokine Res 2000;20:187-90.
11. Khani-Hanjani A, Lacaille D, Hoar D, et al. Association between dinucleotide repeat in non-coding region of interferon-gamma gene and susceptibility to, and severity of, rheumatoid arthritis. Lancet 2000;356:820-5.
12. Goris A, Epplen C, Fiten P, et al. Analysis of an IFN-gamma gene (IFNG) polymorphism in multiple sclerosis in Europe: effect of population structure on association with disease. J Interferon Cytokine Res 1999;19:1037-46.
13. Lee JY, Goldman D, Piliero LM, et al. Interferon-gamma polymorphisms in systemic lupus erythematosus. Genes Immun 2001;2:254-7.
14. Schena FP, Cerullo G, Torres DD, et al. Role of interferon-gamma gene polymorphisms in susceptibility to IgA nephropathy: a family-based association study. Eur J Hum Genet 2006;14:488-96.
15. Rueda B, Martinez A, Lopez-Nevot MA, et al. A functional variant of IFNgamma gene is associated with coeliac disease. Genes Immun 2004;5:517-9.
16. Monsuur AJ, de Bakker PI, Alizadeh BZ, et al. Myosin IXB variant increases the risk of coeliac disease and points toward a primary intestinal barrier defect. Nature genetics 2005;37:1341-4.

17. Cartwright N, Demaine A, Jahromi M, et al. A study of cytokine protein secretion, frequencies of cytokine expressing cells and IFN-G gene polymorphisms in normal individuals. Transplantation 1999;68:1546-52.
18. Nieters A, Brems S, Becker N. Cross-sectional study on cytokine polymorphisms, cytokine production after T-cell stimulation and clinical parameters in a random sample of a German population. Hum Genet 2001;108:241-8.
19. Shnyreva M, Weaver WM, Blanchette M, et al. Evolutionarily conserved sequence elements that positively regulate IFN-gamma expression in T cells. Proceedings of the National Academy of Sciences of the United States of America 2004;101:12622-7.

Supplementary table 1. dbSNP nomenclature, genome location, and sequence of the *IFNG* htSNPs used in this study

htSNP	Location*	Sequence
rs10878760	66816282	TTATGCAAGACTCTTTCTCACCTAG[G/T]TTTGACCTTCCTCCGAGGTTCTCTG
rs2193049	66833189	AGACAGCTTCACAGTAGCCCTTCAT[C/G]TGTTTATACATCAAAGTCTGCATCA
rs2069727	66834490	TCTCCACCTTCCTATTTCCTCCTTC[A/G]TTTCAGAATCTTCCTCTCCCTCATC
rs2069718	66836429	CAAGAGGAAGGTAAATGGTCCACAT[C/T]TTATGAAGCATCATCTAAATGGCCC
rs2069716	66837082	GCATAAAGACAAATCATGTTTTTCA[A/G]AATGTTTTCTAGAAGACAAAGGCCT

* Genome location in bp according to NCBI Build 36.1

Chapter 10

The *SPINK* gene family and coeliac disease susceptibility

MC Wapenaar, AJ Monsuur, J Poell, Ruben van 't Slot, JWR Meijer, GA Meijer, CJJ Mulder, ML Mearin, and C Wijmenga

(Submitted for publication)

ABSTRACT

Objective. The gene family of serine protease inhibitors of the Kazal type (*SPINK*) are functional and positional candidate genes for coeliac disease (CD). Our aim was to assess the gut mucosal gene expression and genetic association of *SPINK1*, *-2*, *-4*, and *–5* in the Dutch CD population.

Methods. Gene expression was determined for all four *SPINK* genes by quantitative reverse-transcription PCR in duodenal biopsy samples from untreated (n=15) and diet-treated patients (n=31), and controls (n=16). Genetic association of the four *SPINK* genes was tested with in total 18 haplotype tagging SNPs, and one coding SNP, in 310 patients and 180 controls. The *SPINK4* study cohort was further expanded to include 479 CD cases and 540 controls. *SPINK4* DNA sequence analysis was performed on six members of a multi-generation CD family to detect possible point mutations or deletions.

Results. *SPINK4* showed differential gene expression, which was at its highest in untreated patients and dropped sharply upon commencement of a gluten-free diet. Genetic association tests for all four *SPINK* genes were negative, including *SPINK4* in the extended case/control cohort. No *SPINK4* mutations or deletions were observed in the multi-generation CD family with linkage to chromosome 9p21-13, nor was the coding SNP disease-specific.

Conclusions. *SPINK4* exhibits CD pathology-related differential gene expression, likely derived from altered goblet cell activity. All four *SPINK* genes tested do not contribute to the genetic risk for CD in the Dutch population.

INTRODUCTION

Coeliac disease (CD) is a chronic inflammatory condition of the small intestine due to an immunological intolerance for the food protein gluten. Patients have to adhere to a life-long diet devoid of gluten to prevent the detrimental effects of a prolonged nutrient and mineral deficiency[1]. Susceptibility for CD is predominantly determined by genetic factors, and the complex inheritance patterns suggest the interaction of multiple genes[2]. It is well established that the adaptive immune response to gluten plays a pivotal role in the pathogenesis of CD. Th1 activation of $CD4^+$ T cells follows gluten-peptide presentation by DQ2 or DQ8 molecules expressed on antigen-presenting cells[3]. The *HLA-DQA* and *-DQB* gene variants coding for these molecules are the major genetic determinants for CD susceptibility[4]. Recently, also the importance of innate immunity in CD pathogenesis was underscored by the observation of induced IL15 expression and NKT cell chemotaxis, through the MICA and NKG2D molecules[5, 6]. However, no genetic contribution of the cognate genes has been demonstrated. The notion of crosstalk between the adaptive and innate immune systems is not limited to CD and gets much attention in studies of the inflammatory process[7]. This raises the

question whether some aspects of innate immunity may contribute to the genetic susceptibility for CD. The innate immune system uses a wide array of defense mechanisms against the invasion of pathogens. These encompass the expression of pattern recognition receptors, release of antimicrobial molecules, and preservation of epithelial barrier and tissue integrity by e.g. serine protease inhibitors[8].

One branch of the family of serine protease inhibitors is that of the Kazal type (SPINK), that originally consisted of four members in humans (*SPINK1, SPINK2, SPINK4*, and *SPINK5*). Recently, as part of a cluster of *SPINK* genes on chromosome 5q32 that already included *SPINK1* and *SPINK5*, five new *SPINK*(-like) members were identified that were located more distally: *SPINK5L2, SPINK6, SPINK5L3, SPINK7*, and *SPINK9*, respectively (NCBI Map Viewer, build 36.1). However, these new members lack functional annotation and were therefore not included in this study. *SPINK* family members 1, 2, and 4 have a comparable size and structure, coded for by 4 exons with a single Kazal type serine protease inhibitor domain. *SPINK5*, in contrast, contains 33 exons that encode 15 inhibitory domains. All four *SPINK* members are thought to be involved in the protection against proteolytic degradation of epithelial and mucosal tissues although their major site of expression may differ. *SPINK1* is expressed in the pancreas and the gastrointestinal tract and mutations in this gene are reported in various forms of pancreatitis[9]. *SPINK2* (located on 4q12) is expressed in testis, epididymis, and seminal vesicle, where its antimicrobial function may be involved in protection of fertility[10]. *SPINK4* was originally isolated from pig intestine[11], and is abundantly expressed in human and porcine goblet cells in the crypts of Lieberkühn, but was also found in monocytes and in the central nervous system[12, 13]. *SPINK5* is expressed in the thymus, vaginal epithelium, Bartolin's glands, oral mucosa, tonsils, and the parathyroid glands[14]. Mutations in *SPINK5* are responsible for Netherton syndrome, a lethal skin disorder characterized by ichthyosis, hair shaft defects, atopy, skin barrier defects, and recurrent bacterial infections[15]. Mouse models of Netherton syndrome have shown enhanced proteolysis of desmoglein 1 and filaggrin in SPINK5 mutants[16, 17]. Moreover, SPINK5 has also been associated with asthma and atopic dermatitis[18]

Interestingly, both *SPINK1* and *SPINK5* are located on chromosome 5q32. This region contains the CELIAC2 susceptibility locus that emerged repeatedly in linkage studies[19]. Despite that this region is rich in candidate cytokine genes and intense mapping efforts were made, no closely associated genes were identified[20]. Likewise, *SPINK4* is located on chromosome 9p13.3 and resides within a linkage region (9p21-13) where we previously identified a novel CD locus that segregated within a four-generation Dutch family[21]. Taken together, the role of *SPINK* genes in epithelial and mucosal protection, and the important genetic locations of *SPINK1*, *SPINK4*, and *SPINK5*, prompted us to subject the four conventional members of the

SPINK family to gene expression and genetic association analyses to ascertain their possible role in CD pathogenesis.

MATERIALS AND METHODS

Patient material

Duodenal biopsy samples were collected by endoscopy as part of a routine CD diagnostic procedure, or to monitor the response to a gluten-free diet in previously diagnosed patients. All patients were classified using the Marsh nomenclature according to the UEGW criteria[22]. Two biopsy samples taken in parallel to those used for histological examination were pooled and used for determination of gene expression. In total, 61 individuals were examined with quantitative reverse-transcription polymerase chain reaction (qRT-PCR), of which 15 were normal controls and 46 were CD patients. The patient group consisted of 15 untreated cases with villous atrophy (MIII), and 31 patients treated with a gluten-free diet who were in various stages of mucosal recovery: MII (crypt hyperplasia; n=11), MI (lymphocyte infiltration; n=8), and M0 (complete remission; n=12). The biopsies of all participating patients were re-evaluated by an experienced pathologist and only CD patients with a proven original Marsh III lesion were included in this study. The genetic association study on all four *SPINK* genes was initially conducted on a cohort of 310 independent CD patients and 180 independent age- and sex-matched controls, all of which were from Dutch Caucasian decent. In a second stage, exclusively focused on all *SPINK4* SNPs, we added 360 controls to a total of 540. In case of three *SPINK4* variants with suggestive P-values, the power of the study was further enhanced by adding 169 extra CD cases to a total of 479. In parallel, we also examined the *SPINK4* gene in a previously described four-generation Dutch CD family[21]. Six family members and a CEPH control were subjected to DNA sequence analysis of the *SPINK4* coding regions and splice sites boundaries. Additionally, we performed *SPINK4* SNP haplotype analysis in this family. All patients and family members that volunteered for this study signed an informed consent. The study was approved by the Medical Ethics Committee of the University Medical Center Utrecht.

Expression study

The isolation of total RNA from biopsy samples and the analysis of gene expression by real-time qRT-PCR on an ABI Prism 7900HT was performed as described before[23]. We used the commercially available Assay-on-Demand test for *SPINK1* (Hs00162154_m1), *SPINK2* (Hs00221653_m1), *SPINK4* (Hs00205508_m1), *SPINK5* (Hs00199260), and the endogenous control gene *GUSB* (PDAR 4326320E) (Applied Biosystems, Foster City, CA). All samples were tested in triplicate on pooled cDNA samples representing each Marsh class. The results were confirmed

with cDNA from individual samples tested in duplicate. Relative levels of gene expression were obtained using the SDS2.1 software (Applied Biosystems).

Genetic association and data analysis

Haplotype tagging SNPs were selected for *SPINK1, SPINK2, SPINK4* and *SPINK5* based on HapMap (Phase I) data using Haploview[24]. For each haploblock containing SNPs in high linkage-disequilibrium, one or more representative SNPs were selected that should capture the genetic variation within that block. For the four *SPINK* genes tested, this resulted in a set of 18 haplotype tagging SNPs and one coding SNP (see table 1 and figure 2). SNP assays were obtained from Applied Biosystems (Foster City, CA) and analyzed on an ABI Prism 7900HT. Hardy-Weinberg Equilibrium (HWE) was evaluated separately in cases and controls, for all SNPs tested. Allele frequencies were compared between cases and controls and P-values were obtained by χ^2 analysis.

DNA sequence analysis

DNA sequence analysis was performed on *SPINK4* in six members of a four-generation Dutch CD family, and one CEPH control (family 1331, individual 02). Of these six family members, four were affected (index 02, 08, 32, and 41) and carried the disease-linked haplotype, and two were non-affected (index 21 and 31) without this haplotype (figure 3). All coding sequences of the *SPINK4* gene were PCR-amplified, including the intron-exon boundaries (for primers and protocols see supplementary data table 1). PCR products were examined on a 2% agarose gel and purified with a Millipore Vacuum Manifold (Billerica, MA). Samples were prepared with the BigDye terminator cycle sequencing ready kit (Applied Biosystems, Foster City, CA) according to the manufacturer's protocol. PCR and sequencing amplification were performed on a GeneAmp PCR system 9700 (Perkin Elmer, Foster City, CA). Sequences were run on an ABI Prism 3730 analyzer (Applied Biosystems, Foster City, CA). Analysis and sequence alignment was carried out with Sequence Navigator (Applied Biosystems, Foster City, CA) and Vector NTI (InforMax Inc, MA, USA) software packages.

RESULTS

SPINK gene expression in the CD mucosa

The expression of all four conventional members of the human *SPINK* family was determined by real-time qRT-PCR on duodenal biopsy-derived cDNA pools from normal controls and CD patients, either untreated or in various stages of remission on a gluten-free diet. The results shown in figure 1 indicate that only *SPINK4* (figure 1C) is differentially expressed and that its transcriptional activity, which is at its highest in Marsh III (20-fold compared to controls), decreases sharply

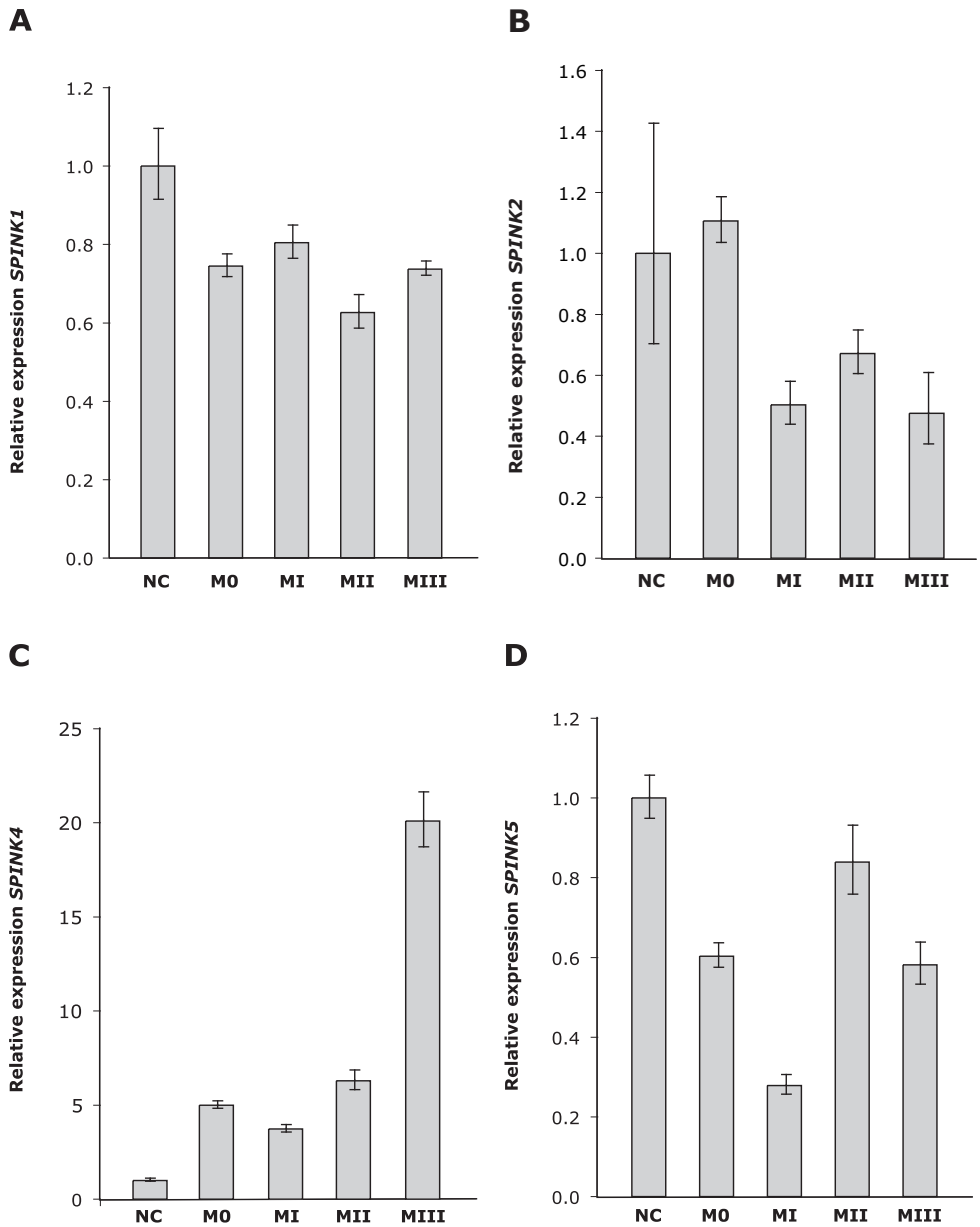
(four-fold) when patients improve and make a transition to Marsh II. To preclude that the results for *SPINK4* might be biased by fortuitous differences in individual expression levels within the generated pools, we also examined the control and case samples each separately. This did not change the observed drop in *SPINK4* expression during tissue recovery (see supplementary figure 1). Likewise, we performed the same analysis for the other three *SPINK* genes without affecting the profile already observed in the pools (results not shown). We also examined the relative expression of the four *SPINK* genes with respect to each other in the normal intestinal mucosa. This showed that both *SPINK1* and *SPINK4* have the highest expression, which is respectively 480-fold and 240-fold higher compared to *SPINK2*, while *SPINK5* is in the same order of magnitude (five-fold) as *SPINK2* (figure 1E). In conclusion, only the *SPINK4* gene appears to be differentially regulated in the intestinal mucosa during recovery from the gluten-evoked CD lesion. This observation prompted us to examine whether *SPINK4*, or any of the other *SPINK* genes, could also be causally related to the CD pathogenesis.

Genetic association analysis of SPINK genes

We designed a haplotype tagging SNP strategy to capture all genetic variation in *SPINK1*, *-2*, *-4*, and *–5*. An overview of these four *SPINK* genes with their genomic organization, linkage-disequilibrium structure, and the position of the haplotype tagging SNPs used, is depicted in figure 2. Initially these haplotype tagging SNPs were tested in 310 CD cases and 180 controls (table 1) and showed no significant association for any of the haplotype tagging SNPs in the four SPINK genes. Despite the initial negative result we decided to pursuit *SPINK4* further because it is expressed in goblet cells[12], displayed a CD pathology-related differential expression in the intestinal mucosa, and maps within a CD linkage region[21]. Initially, we expanded the control group with 360 samples to a total of 540 for all SPINK4 SNPs tested. As a result the Val7Ile coding variant rs706107 and its flanking haplotype tagging SNPs rs891671 and rs706109 yielded suggestive, but non-significant, P-values of 0.0595, 0.0510, and 0.1122, respectively (data not shown). To increase the power of the study even further we subsequently added 169 CD cases to a total of 479. The effect on the P-values of the three SNPs tested was such that they dropped below the significance threshold (see table 2). From this we conclude that the four *SPINK* genes tested do not contribute to the genetic susceptibility in the Dutch CD population.

SPINK4 sequence analysis in a multi-generation family

We have previously described a four-generation CD family with an extraordinary high incidence of affected individuals (see figure 3). The disease segregated with a grand-maternal haplotype on chromosome 9p21-13[21], a region that encompasses


A
Relative expression SPINK1
1.2
1.0
0.8
0.6
0.4
0.2
0.0
NC
M0
MI
MII
MIII
B
Relative expression SPINK2
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
NC
M0
MI
MII
MIII
C
Relative expression SPINK4
25
20
15
10
5
0
NC
M0
MI
MII
MIII
D
Relative expression SPINK5
1.2
1.0
0.8
0.6
0.4
0.2
0.0
NC
M0
MI
MII
MIII

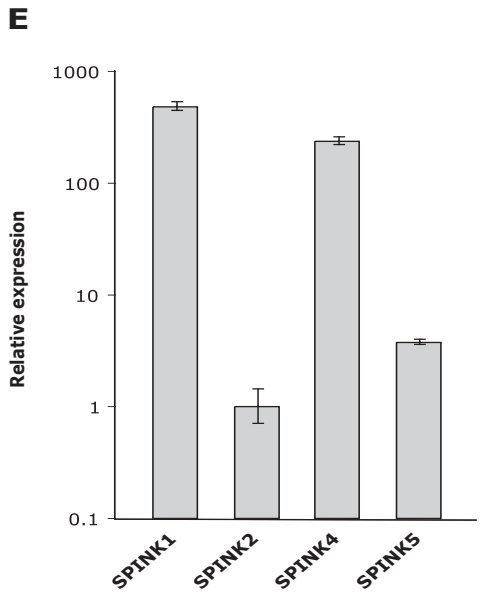
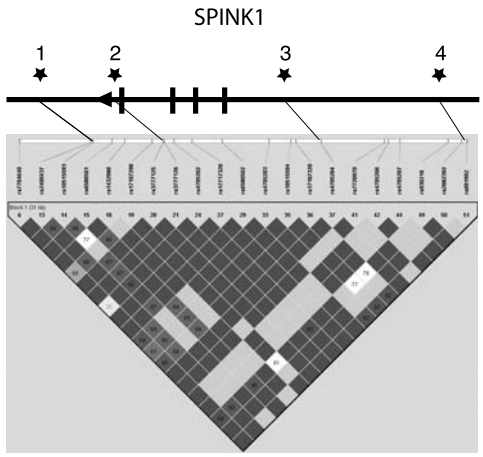
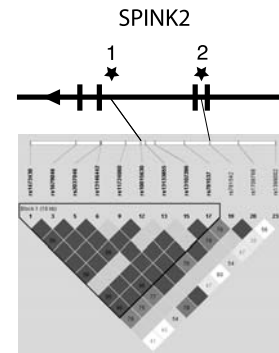
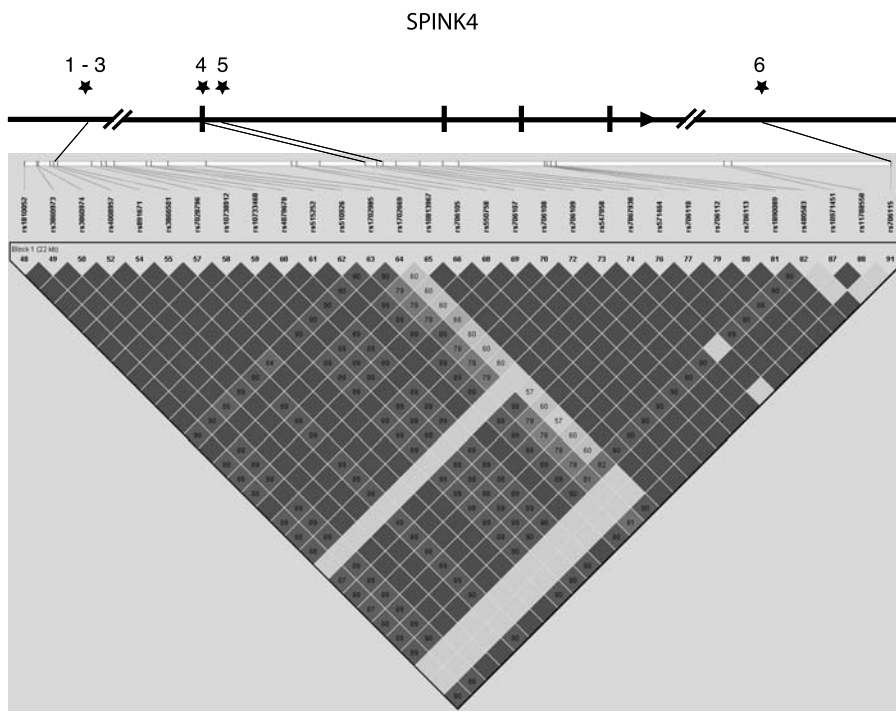


Figure 1. Results of qRT-PCR of *SPINK* genes in normal controls (NC) and CD patients, either untreated (MIII), or on a gluten-free diet (MII-M0). The Marsh (M) stages refer to the pathological conditions of the mucosa, characterized by atrophy of the villi (MIII); hyperplastic crypts between the villi (MIII - MII); and enhanced lymphocyte infiltration (MIII – MII - MI). Stage M0 indicates complete remission, comparable to controls. The genes tested were: *SPINK1* (**A**); *SPINK2* (**B**); *SPINK4* (**C**); and *SPINK5* (**D**). Measurements were made in triplicate, on pools of separately prepared cDNA samples. Expression data were normalized to the normal control pool. (**E**) Relative expression of all four *SPINK* genes with respect to *SPINK2* in the healthy duodenal mucosa. Note the logarithmic scale here. The *GUSB* gene was used as endogenous control in all tests. Error bars indicate standard deviations.


SPINK1
1
2
3
4


SPINK2
1
2


SPINK4
1 - 3
4 5
6

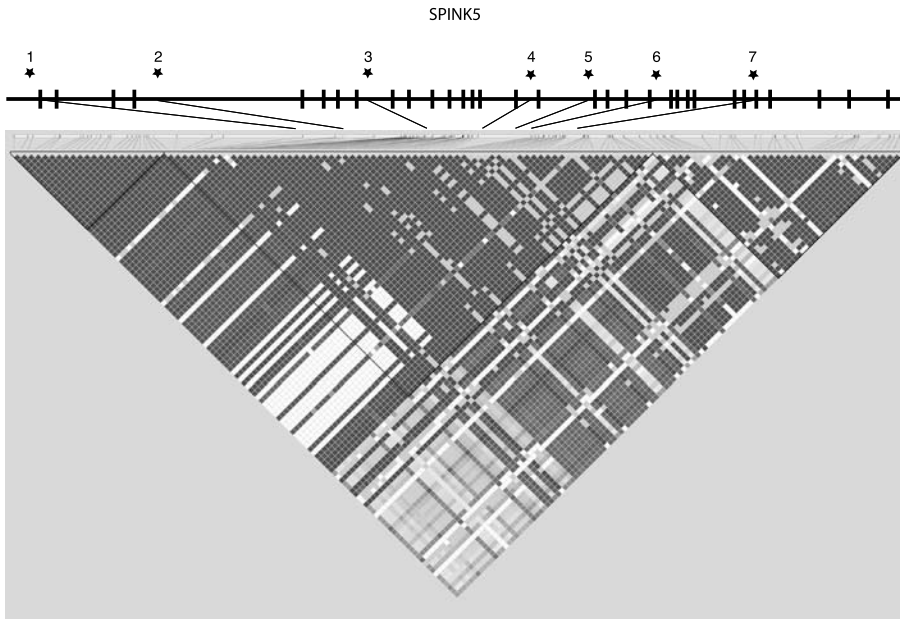


Figure 2. Genomic organization of the four *SPINK* genes. The upper horizontal line indicates exon locations (vertical bars) and SNP positions (numbered asterisks). The SNPs are numbered for each gene consecutively as they appear in table 1. *SPINK4* SNP number 4 represents the non-synonymous (Val7Ile) coding SNP rs706107. The arrow points indicate the orientation of transcription. The lower portion of the figure shows the pairwise linkage-disequilibrium structure between indicated SNPs given by D′ statistics based on the European population in the HapMap database (Phase II). Darker red intensities indicate higher D′-values (numbers indicate D′-value, while SNP pairs without number have a D′=1).

SPINK4. The apparent dominant inheritance pattern could be caused by a mutation that is rare in the general CD population, but present with a high phenotypic penetration in this specific family. To assess if any functional variants of the *SPINK4* gene were present in this family we sequenced all its exons and intron-exon boundaries in six family members. However, we did not observe mutations in any of the samples tested (results not shown). Neither was the exon 1 coding SNP rs706107 specific for affected individuals since all seven individuals tested (including the CEPH control) carried the most frequent GG genotype (figure 3). To exclude the possibility of deletions in *SPINK4* to be misinterpreted from the sequence data as homozygous genotypes, we also performed segregation analysis of the grand-parental *SPINK4* haplotypes within the entire family but observed no suspect inheritance pattern (figure 3). In conclusion, we have found no evidence that *SPINK4* is a candidate gene for the chromosome 9p21-13 CD locus in the Dutch population in general, or in the multi-generation Dutch CD family specifically.

Table 1. Allelic distribution of *SPINK* haplotype tagging SNPs in a Dutch CD case-control cohort

				Cases (n=310)				Controls (n=180)					
			Allele	Allele counts		Allele freq (%)		Allele counts		Allele freq (%)			
Gene name	SNP i.d.	Position*	Minor/ Major	Minor	Major	Minor	Major	Minor	Major	Minor	Major	χ^2	P-value
SPINK1	rs10515593	147,178,993	A/G	119	449	21	79	76	268	22.1	77.9	0.166	0.6834
SPINK1	rs3777125	147,184,010	C/G	228	336	40.4	59.6	150	200	42.9	57.1	0.527	0.4681
SPINK1	rs4705204	147,195,313	C/A	125	429	22.6	77.4	60	276	17.9	82.1	2.813	0.0935
SPINK1	rs891992	147,205,707	G/A	77	483	13.8	86.2	47	303	13.4	86.6	0.019	0.8906
SPINK2	rs10015630	57,520,070	A/G	216	328	39.7	60.3	136	216	38.6	61.4	0.102	0.7489
SPINK2	rs781542	57,528,232	G/A	220	390	36.1	63.9	123	221	35.8	64.2	0.009	0.9237
SPINK4	rs563353	33,206,428	G/C	120	472	20.3	79.7	67	279	19.4	80.6	0.112	0.7375
SPINK4	rs563512	33,206,484	G/A	50	556	8.3	91.7	36	310	10.4	89.6	1.243	0.2648
SPINK4	rs891671	33,222,271	G/T	70	540	11.5	88.5	52	304	14.6	85.4	1.998	0.1575
SPINK4	rs706107**	33,230,225	A/G	117	499	19	81	81	269	23.1	76.9	2.358	0.1246
SPINK4	rs706109	33,230,668	A/G	118	486	19.5	80.5	82	276	22.9	77.1	1.549	0.2133
SPINK4	rs706115	33,243,605	G/C	77	537	12.5	87.5	52	308	14.4	85.6	0.716	0.3975
SPINK5	rs3756688	147,422,972	G/A	232	380	37.9	62.1	135	221	37.9	62.1	0.000	0.9968
SPINK5	rs4472254	147,433,830	A/C	240	376	39	61	142	214	39.9	60.1	0.081	0.7757
SPINK5	rs4519913	147,452,004	G/A	279	331	45.7	54.3	168	192	46.7	53.3	0.079	0.7792
SPINK5	rs1422987	147,466,552	T/C	49	559	8.1	91.9	27	329	7.6	92.4	0.070	0.7917
SPINK5	rs3815740	147,471,386	G/A	38	574	6.2	93.8	21	337	5.9	94.1	0.047	0.8291
SPINK5	rs2052532	147,476,500	G/A	200	414	32.6	67.4	117	241	32.7	67.3	0.001	0.9723
SPINK5	rs3764930	147,485,309	G/A	199	413	32.5	67.5	117	243	32.5	67.5	0.000	0.9958

*Basepair position according to NCBI build 35.1
**Coding SNP, non-synonymous change (Val7Ile)

Table 2. Allelic distribution of three selected *SPINK4* SNPs in the extended Dutch CD case-control cohort

				Cases (n=479)				Controls (n=540)					
			Allele	Allele counts		Allele freq (%)		Allele counts		Allele freq (%)			
Gene name	SNP i.d.	Position*	Minor/ Major	Minor	Major	Minor	Major	Minor	Major	Minor	Major	χ^2	P-value
SPINK4	rs891671	33,222,271	G/T	126	814	13.4	86.6	158	904	14.9	85.1	0.889	0.3457
SPINK4	rs706107**	33,230,225	A/G	205	737	21.8	78.2	242	814	22.9	77.1	0.382	0.5365
SPINK4	rs706109	33,230,668	A/G	212	724	22.6	77.4	242	816	22.9	77.1	0.014	0.9053

*Basepair position according to NCBI build 35.1
**Coding SNP, non-synonymous change (Val7Ile)

DISCUSSION

Chronic inflammatory conditions and autoimmune disorders are typically characterized by a deregulated adaptive and innate immune system. The innate defense consists of multiple components that include physical barriers, antimicrobial molecules, pattern recognition receptors, circulating phagocytes, and the complement system[7]. A breach of the epithelial barrier and loss of microbial containment is often the first of a series of events that trigger or sustain chronic inflammatory diseases[25], as described e.g. in Crohn's disease, atopic eczema, asthma, and psoriasis[26]. In CD the gut-lumen separation is undermined by dietary gluten that evokes a combined innate and adaptive immune response[27]. It is the joined action of gluten peptides, environmental factors, and genetic determinants that precipitates this enteropathy. The *HLA* locus is the major genetic contribution to the adaptive Th1 reaction[4]. Recently, we identified *MYO9B* as a susceptibility gene in the Dutch population that possibly has an effect on epithelial barrier integrity[28]. Several other studies have underscored the involvement of innate immunity in CD, however without identification of underlying causative gene variants[27]. Interestingly, it was also reported that the epithelial glycocalyx and the bacterial composition in the CD gut is distinct[29, 30].

In search of genes that may have a primary contribution to CD pathogenesis we focused our attention to the *SPINK* family of serine protease inhibitors that play an important role in tissue preservation through containment of uncontrolled proteolysis and bacterial growth. In this study we demonstrate differential gene expression of mucosal *SPINK4* in CD. Crypt hyperplasia is a feature of the Marsh III and Marsh II stages of CD and the concomitant increase of the number of goblet cells may contribute to the increased *SPINK4* expression. However, the observed sharp decrease in gene expression sets in during the MIII/MII transition while crypt normalization is observed only later at the MII/MI recovery phase. This suggests that *SPINK4* down-regulation sets in soon after commencement of the gluten-free diet. This *SPINK4* differential expression probably reflects altered goblet cell activity, but its functional significance and regulatory mechanism in CD pathology remains to be established.

The combination of functional relevance and mapping to CD linkage intervals pointed to the *SPINK* family members as attractive functional and positional candidate genes. We have chosen a robust strategy for genetic association testing based on haplotype tagging SNPs and linkage-disequilibrium structure of the *SPINK* loci, applied to a considerably sized Dutch case-control cohort. With our study design we had 75% power to detect association with *SPINK1*, *-2*, and *–5* (relative risk 2.0; allele frequency 0.1 – 0.45; 95% confidence interval), while these was even 95% (RR 2.0), and 80% (RR 1.6) for *SPINK4*. In parallel, we examined the extended Dutch CD family for variants and deletions in *SPINK4*. We hypothesized

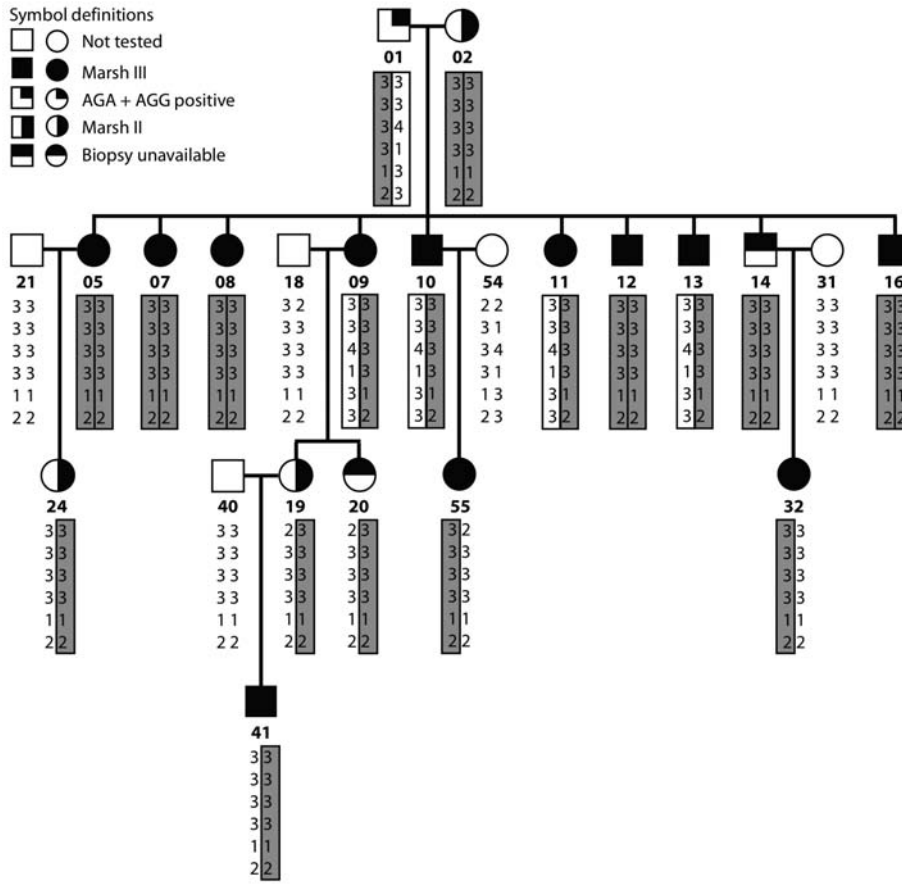


Figure 3. Pedigree of the Dutch multi-generation CD family. Only affected descendents are depicted (ten out of thirteen siblings in the second generation were affected). The grand-parental *SPINK4* haplotypes are boxed, and shaded where identical to the grand-maternal at-risk haplotype (non-informative). The SNPs are ordered (top-to-bottom) as they appear in table 1. Genotype numbers 1, 2, 3, and 4 refer to A, C, G, and T alleles, respectively. Sequence analysis was performed on family members 02, 08, 21, 31, 32, and 41. Family member index numbers are indicated in bold.

that a specific *SPINK4* mutation, though rare in the general population, could have a dramatic impact on mucus composition, bacterial containment, and gluten-sensitivity, thereby explaining the apparent dominant and high penetration inheritance pattern in our extended CD family. With both approaches we were not able to establish a genetic involvement of the *SPINK* genes tested. However, we cannot completely rule out the possibility of a rare non-coding mutation in *SPINK4* (outside the splice donor and acceptor regions) that might specifically segregate in this atypical CD family, characterized by an exceptional high prevalence of affected members.

Despite this negative result in the Dutch CD population, we cannot formally rule out the possibility of genetic contribution of *SPINK* genes to CD in other European populations like the Italian in whom, unlike the Dutch[31], chromosome 5q linkage was established[32, 33]. Genuine population heterogeneity has been reported before, e.g. between *CARD15/NOD2* and Crohn's disease[34, 35], and between *SPINK5* and asthma[18, 36]. The new *SPINK* members on chromosome 5q (*SPINK5L2, SPINK6, SPINK5L3, SPINK7*, and *SPINK9*) were not part of this study. Currently, no functional annotation is available for these genes that are located near *SPINK1* and *SPINK5* in a chromosomal region that appears to have been subjected to gene duplication during evolution. Therefore, we cannot exclude their possible involvement in coeliac disease or any other inflammatory disorder. Finally, the study presented here, based on high-throughput extraction of genetic information from a family of functional-positional candidate genes, could equally well be extended to scrutinize entire molecular pathways for causative genes and mechanisms of pathogenesis.

ACKNOWLEDGEMENTS

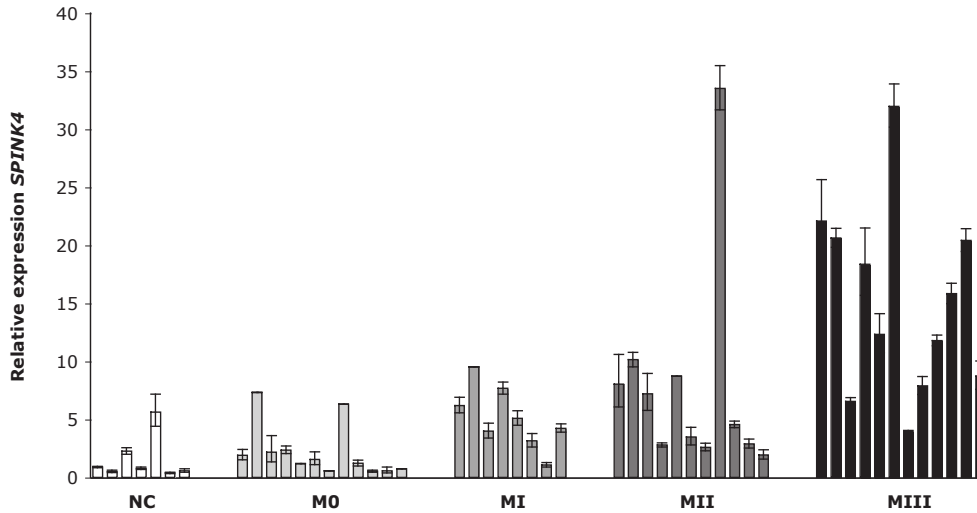
We wish to thank the Dutch coeliac disease patients and their family for participating in this study. We like to thank Dr. Joachim Schweizer (LUMC) for biopsy sampling. This study was supported by grants from the Dutch Digestive Diseases Foundation (MLDS, WS00-13) and the Netherlands Organization for Scientific Research (ZonMW, 912-02-028). This research was carried out within the Celiac Disease Consortium, an innovative cluster approved by the Netherlands Genomics Initiative and partially funded by a Dutch government grant (BSIK03009).

REFERENCES

1. Green PH, Jabri B. Coeliac disease. Lancet 2003;362:383-91.
2. van Heel DA, Hunt K, Greco L, et al. Genetics in coeliac disease. Best Pract Res Clin Gastroenterol 2005;19:323-39.
3. Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. Nat Rev Immunol 2002;2:647-55.
4. Koning F, Schuppan D, Cerf-Bensussan N, et al. Pathomechanisms in coeliac disease. Best Pract Res Clin Gastroenterol 2005;19:373-87.
5. Hue S, Mention JJ, Monteiro RC, et al. A direct role for NKG2D/MICA interaction in villous atrophy during coeliac disease. Immunity 2004;21:367-77.
6. Meresse B, Chen Z, Ciszewski C, et al. Coordinated induction by IL15 of a TCR-independent NKG2D signaling pathway converts CTL into lymphokine-activated killer cells in coeliac disease. Immunity 2004;21:357-66.
7. Hoebe K, Janssen E, Beutler B. The interface between innate and adaptive immunity. Nat Immunol 2004;5:971-4.

8. Kimbrell DA, Beutler B. The evolution and genetics of innate immunity. Nat Rev Genet 2001;2:256-67.
9. Pfutzer RH, Whitcomb DC. SPINK1 mutations are associated with multiple phenotypes. Pancreatology 2001;1:457-60.
10. Rockett JC, Patrizio P, Schmid JE, et al. Gene expression patterns associated with infertility in humans and rodent models. Mutat Res 2004;549:225-40.
11. Agerberth B, Soderling-Barros J, Jornvall H, et al. Isolation and characterization of a 60-residue intestinal peptide structurally related to the pancreatic secretory type of trypsin inhibitor: influence on insulin secretion. Proceedings of the National Academy of Sciences of the United States of America 1989;86:8590-4.
12. Metsis M, Cintra A, Solfrini V, et al. Molecular cloning of PEC-60 and expression of its mRNA and peptide in the gastrointestinal tract and immune system. J Biol Chem 1992;267:19829-32.
13. Norberg A, Gruber S, Angelucci F, et al. Identification of the bioactive peptide PEC-60 in brain. Cell Mol Life Sci 2003;60:378-81.
14. Magert HJ, Standker L, Kreutzmann P, et al. LEKTI, a novel 15-domain type of human serine proteinase inhibitor. J Biol Chem 1999;274:21499-502.
15. Bitoun E, Chavanas S, Irvine AD, et al. Netherton syndrome: disease expression and spectrum of SPINK5 mutations in 21 families. J Invest Dermatol 2002;118:352-61.
16. Descargues P, Deraison C, Bonnart C, et al. Spink5-deficient mice mimic Netherton syndrome through degradation of desmoglein 1 by epidermal protease hyperactivity. Nature genetics 2005;37:56-65.
17. Hewett DR, Simons AL, Mangan NE, et al. Lethal, neonatal ichthyosis with increased proteolytic processing of filaggrin in a mouse model of Netherton syndrome. Hum Mol Genet 2005;14:335-46.
18. Blumenthal MN. The role of genetics in the development of asthma and atopy. Curr Opin Allergy Clin Immunol 2005;5:141-5.
19. Babron MC, Nilsson S, Adamovic S, et al. Meta and pooled analysis of European coeliac disease data. Eur J Hum Genet 2003;11:828-34.
20. Ryan AW, Thornton JM, Brophy K, et al. Chromosome 5q candidate genes in coeliac disease: genetic variation at IL4, IL5, IL9, IL13, IL17B and NR3C1. Tissue Antigens 2005;65:150-5.
21. van Belzen MJ, Vrolijk MM, Meijer JW, et al. A genomewide screen in a four-generation Dutch family with coeliac disease: evidence for linkage to chromosomes 6 and 9. The American journal of gastroenterology 2004;99:466-71.
22. Report of a working group of the United European Gastroenterology Week in Amsterdam. When is a coeliac a coeliac? Eur J Gastroenterol Hepatol 2001;13:1123-28.
23. Wapenaar MC, van Belzen MJ, Fransen JH, et al. The interferon gamma gene in coeliac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility. J Autoimmun 2004;23:183-90.
24. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.

25. Tlaskalova-Hogenova H, Stepankova R, Hudcovic T, et al. Commensal bacteria (normal microflora), mucosal immunity and chronic inflammatory and autoimmune diseases. Immunol Lett 2004;93:97-108.
26. Schreiber S, Rosenstiel P, Albrecht M, et al. Genetics of Crohn disease, an archetypal inflammatory barrier disease. Nat Rev Genet 2005;6:376-88.
27. Londei M, Ciacci C, Ricciardelli I, et al. Gliadin as a stimulator of innate responses in coeliac disease. Mol Immunol 2005;42:913-8.
28. Monsuur AJ, de Bakker PI, Alizadeh BZ, et al. Myosin IXB variant increases the risk of coeliac disease and points toward a primary intestinal barrier defect. Nature genetics 2005;37:1341-4.
29. Forsberg G, Fahlgren A, Horstedt P, et al. Presence of bacteria and innate immunity of intestinal epithelium in childhood coeliac disease. The American journal of gastroenterology 2004;99:894-904.
30. Tjellstrom B, Stenhammar L, Hogberg L, et al. Gut microflora associated characteristics in children with coeliac disease. The American journal of gastroenterology 2005;100:2784-8.
31. Van Belzen MJ, Meijer JW, Sandkuijl LA, et al. A major non-HLA locus in coeliac disease maps to chromosome 19. Gastroenterology 2003;125:1032-41.
32. Greco L, Corazza G, Babron MC, et al. Genome search in coeliac disease. Am J Hum Genet 1998;62:669-75.
33. Percopo S, Babron MC, Whalen M, et al. Saturation of the 5q31-q33 candidate region for coeliac disease. Ann Hum Genet 2003;67:265-8.
34. Lesage S, Zouali H, Cezard JP, et al. CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. American journal of human genetics 2002;70:845-57.
35. Croucher PJ, Mascheretti S, Hampe J, et al. Haplotype structure and association to Crohn's disease of CARD15 mutations in two ethnically divergent populations. Eur J Hum Genet 2003;11:6-16.
36. Jongepier H, Koppelman GH, Nolte IM, et al. Polymorphisms in SPINK5 are not associated with asthma in a Dutch population. J Allergy Clin Immunol 2005;115:486-92.



Supplementary figure 1. Relative expression of *SPINK4* in normal controls (NC) and CD patients, either untreated (MIII), or on a gluten-free diet (MII-M0). Measurements were made in duplicate by qRT-PCR. All expression data were normalized to the averaged value of the seven normal controls. The *GUSB* gene was used as endogenous control in all tests. Error bars indicate standard deviations.

Supplementary table 1. Primers and reaction conditions used for *SPINK4* DNA sequence analysis

Exon	Product length (bp)	5'-primer	3'-primer	Input genomic DNA (ng)
1	287	TAAATAGGCCCCCATCCTTC	CAGAACTTTCCCCAGCACAG	100
2	222	GCAGAAGCAGTCCTCAGACC	AAGGTGTCTGAGCAGGATCG	100
3	382	TCTTCACTGGACCCACACTG	GGGGCATATGTTGGAGGAC	100
4	278	GATGTCTGGATGGACTGTGC	CTCCTATCCCCTCTGCAACC	100

Reactions were carried out in 20 ul volume containing 2 ul dNTPs (0.5mM), 0.3 ul of a homemade Taq polymerase, 1.2 ul of 25 mM MgCl2, 100 ng of each oligonucleotide primer, and 2 ul 10X PCR buffer. Cycling consisted of denaturation at 94°C for 7 minutes, followed by 32 cycles at 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 1 minute. The final extension step was at 72°C for 4 minutes.

GENERAL DISCUSSION

The genomics approach, an evaluation

In this thesis we describe a genomics approach – an integration of genetics and gene expression profiling – taken to enhance our understanding of the pathogenesis of coeliac disease. Our initial intention was to compare genome-wide linkage data with global gene expression profiles to identify candidate genes for genetic association studies (chapters 1 and 2). By pinpointing differentially expressed genes from within the linkage intervals, typically comprising hundreds of genes, we hoped to expedite the search for positional candidate genes, facilitated by bioinformatics tools like TEAM[1]. This approach was based on two premises: 1) mutated or variant genes are differentially expressed; and 2) a sufficient number of linkage intervals are available to allow a logically connection by differentially expressed genes belonging to a single or limited number of molecular pathways. Although the concept is in principle useful, it was not applicable to our study on coeliac disease: 1) genes that were differentially expressed were not genetically associated; 2) genetically associated genes were not differentially expressed; and 3) linkage intervals in coeliac disease are generally too few in number, and too large in size, to warrant success. The inter-relationship between genetic variants and their effect on transcriptional levels of genes in a single pathway is probably far more complex and will require more sophisticated bioinformatics approaches than we had envisaged initially (see also the penultimate paragraph).

If the genomics approach did not meet our initial expectations, was there in fact any additional value to this strategy? After having read the previous chapters you will not be surprised that the answer is affirmative. The identification of *MYO9B* by linkage analysis and fine-mapping by genetic association could only be fully appreciated in a functional context[2], after more insight was gained into intestinal permeability and epithelial differentiation, inspired by our expression profiling studies (chapter 4). Moreover, recognition of the potential role of *MYO9B* in permeability encouraged us to conduct a genomics study of the entire tight-junction gene network (section III). This yielded two more permeability-related candidate genes and much new information on the normal and pathological activity of this pathway in the duodenum. The notion that both coeliac disease and inflammatory bowel disease are marked by an impaired intestinal barrier led us to successfully test all three permeability genes in the latter disorder[3] (chapter 7).

The observations from our first microarray study, pointing towards the role of differentiation and proliferation (chapter 4), shaped our thinking on coeliac disease pathogenesis and paved the way for a far more elaborate and sophisticated expression study (chapter 5). This study compellingly demonstrated that the gluten-provoked inflammation induced arrested enterocyte differentiation, directly related to the diverse clinical symptoms seen in coeliac patients. In conclusion, the integration of genetics and gene expression profiling within a single genomics

approach helped to lay a solid foundation for the successful dissection of the coeliac disease pathogenesis.

The epithelial trinity: differentiation, barrier maintenance, and polarity

The endoderm, in contrast to the stratified epidermis, consists of a single layer of epithelial cells, but it is still capable of performing many vital functions indispensable to its host. It provides the absorption of essential nutrients through the expression of transporters and metabolic enzymes. Specialized absorptive cells, the enterocytes, have an apical structure of microvilli (the brushborder) that facilitate this absorptive capacity by surface enlargement and transporter enrichment. Because of its close contact with the intestinal lumen and its contents, the epithelium is also strongly involved in the defense against harmful chemical agents and microbial invaders. For this purpose the epithelium is equipped with a number of defensive lines. Paneth cells at the bottom of the crypts produce antibiotic peptides: defensins and cathelicidins[4, 5]. Goblet cells along the crypt-villus axis produce mucins, which are also endowed with anti-bacterial characteristics[6]. Apical junctions between the epithelial cells, the tight junction and the adherence junction, prevent the influx of pathogens along the paracellular route into the lamina propria and the circulation[7]. The expression of a number of phase I and phase II detoxification genes like ABC transporters, cytochome P450 mono-oxygenases, and conjugating enzymes, neutralize or excrete xeno- and endobiotics before they become harmful to the enterocyte or the organism as a whole[8, 9]. This multitude of specialized functions requires epithelial cells to be fully differentiated. A small number of stem cells at the bottom of the crypts, just above the Paneth cells, produce a regular stream of new epithelial cells. These migrate either to the bottom to differentiate into Paneth cells, or to the top to differentiate into enterocytes, goblet cells, and enteroendocrine cells[10]. Differentiation of enterocytes progresses as they migrate along the crypt-villus axis until they are shed from the villus tip into the lumen[11]. Inflammation of the gut, as in the case of the gluten-precipitated interferon-gamma release, impedes this differentiation process and results in multiple functional defects associated with the clinical features of coeliac disease (chapters 3 and 5).

The paracellular route is safeguarded by the junctional complexes, of which the most apical one, the tight junction, is most important. The function of the tight junction is to: 1) control a selective paracellular flux of ions and macromolecules (intestinal permeability); 2) maintain polarity by forming a diffusion barrier between apical and basolateral transmembrane proteins; and 3) prevent entrance of pathogens and commensals along the paracellular route[12]. Tight junctions are formed by the transmembrane proteins occludin, junction adhesion molecules, and claudins[13]. Particularly the large claudin family provides differential selectivity

for ion transport in various tissues[14]. In this thesis (chapter 6) we show that the transcriptional regulation of particularly these claudins in coeliac disease patients show extreme differences, possibly related to the enhanced intestinal permeability ('leaky gut'), so typical for this disorder. We also demonstrated that three pairs of claudins are co-regulated: in two pairs coinciding with a genomic juxtaposition, and in another pair probably due to functional constraints. We also provided, for the first time, a transcriptome of the tight junction gene network that included, in addition to transmembrane proteins, adaptors, signal transduction molecules, transcriptional regulators, and cytoskeletal proteins (chapter 6). From this it became apparent that the claudin composition in the small intestine, and thus its molecular-influx selectivity, has been preserved from the time that the ancestors of mice and man diverged, some 75 million years ago[15-17].

In this study we also observed high-level expression of *RHOA*, and cytoskeletal proteins. The small GTPase RhoA is involved in many biological processes including migration, differentiation, growth, and control of tight junctions[18, 19]. RhoA drives these processes through the regulation of the actinomyosin cytoskeleton[20, 21]. Activated RhoA leads to contraction of the actin fibers, thereby preventing the simultaneous formation of junctional complexes and cellular polarity[22]. Activation of RhoA has a weakening effect on the existing tight junctions by strengthening the forces pulling on them, resulting in enhanced permeability[23]. In this context it is worth mentioning the role of the gene *MYO9B*. Myosin IXB is an unconventional myosin that contains a RhoGAP domain that is probably involved in the inactivation of RhoA[2, 24]. A proper *MYO9B* activity may thus prove to be essential for a functional tight junction, and control of permeability. Previously we reported genetic association of *MYO9B* with coeliac disease that was confined to the 3'-region of the gene that contains the RhoGAP domain[2]. A mutation in this domain might result in constitutively over-activated RhoA with an effect on the tight junction, cellular polarity, migration, and even differentiation. These are also the hallmarks of coeliac disease pathology.

The relation between the tight junction, differentiation, and growth, goes beyond RhoA control. An essential constituent of the tight junction complex is the signaling cascade between the membrane and nucleus[23]. This intricate signaling network was also the subject of our transcriptome (chapter 6) and genetic association study (chapter 7). It has been reported that a flow of information travels between junctional complexes and the nucleus to control growth and differentiation[19, 23]. Tight junction proteins act as cell-contact sensors, promoting growth inhibition and cellular differentiation.

Various leukocyte cell types also express tight junction proteins on their membrane[25-28]. This allows them to fuse with the tight junction and facilitate an efficient passage across the epithelial barrier. This has been described for the

extrusions of dendritic cells that sample antigen from the intestinal lumen for presentation to T-cells [29]. In this respect the recent report that dendritic cells are the major antigen-presenting cells in the coeliac mucosal lesion is relevant[30]. The RhoA controlling activity of *MYO9B* and its association to coeliac disease might not only be relevant to the epithelial cells and the tight junction, but might also affect other cell types characterized by a dynamic cytoskeleton like neutrophils (diapedesis), T-cells (immune synapse), and dendritic cells (dendrites).

Another intriguing observation was the activation of the growth factor gene *EGF* in untreated coeliac disease, which persisted in patients in remission (chapter 6). It has been amply documented that EGF protects the integrity of the epithelium and the tight junction[31, 32]. The *EGF* activation we observed may thus be a response of the intestinal mucosa to deal with the gluten-induced epithelial damage or restoration of the barrier function. In newborn rats, induction of a 'flat' and 'leaky' gut by combined interferon-gamma and gluten presentation can be prevented by feeding rat mother milk or EGF-supplemented formula milk, but not by formula milk alone[33]. Human newborns have a natural 'leaky' gut that persists for a relatively long time compared to other mammals[34]. Epidemiological studies have suggested the protective effect of breast-feeding probably due to the multiple growth factors contained in breast milk[35, 36].

In summary, epithelial differentiation, tight junction integrity, and maintenance of polarity are tightly integrated processes, partly controlled by the actinomyosin cytoskeleton with a prominent role for RhoA, and preserved by the action of EGF.

Coeliac disease and inflammatory bowel disease: is there more than meets the eye?

Coeliac disease and inflammatory bowel disease (IBD) each seem to have attracted their own community of researchers. Although coeliac disease is about ten-fold more frequent in Western populations (~1% prevalence) than IBD, it is probably the more dramatic course of IBD, often requiring surgical intervention, that has made the latter draw more attention. But is this separation of the minds really so natural, or is it rather a result of a historical anomaly. I would argue for the latter. Admittedly, a gluten-induced absorption disorder of the small intestine does seem radically different to a spontaneous inflammation of the colon due to bacterial overgrowth. But since we are in general talking about the same organ, the intestinal tract, it should not come as a surprise that the two parts (the small intestine and the colon) may interfere with each other.

Studies on the co-occurrence of the two disorders show that the frequency of IBD, comprised of Crohn's disease and ulcerative colitis, is elevated in coeliac disease patients[37-40], which may suggest that they do share at least some factors

in their pathogenesis. Genetic linkage studies show that coeliac disease and IBD share a common susceptibility region on chromosome 5q31-q33 and 19p13[41]. In this thesis we also demonstrated that two genes from the tight junction complex are genetically associated with both disorders. These genes are also involved in the formation and maintenance of cellular polarity. Interestingly, *DLG5*, a gene associated with IBD is also involved in polarity[42, 43]. This may suggest that similar pathways, although not necessarily identical genes, could be involved in both disorders. We have recently established genetic association of *MYO9B* with coeliac disease[2] and IBD[3], particularly ulcerative colitis. Also *MYO9B* is involved in the regulation of tight junction-mediated permeability through its effect on the actino-myosin cytoskeleton via the small GTPase RhoA[19, 24]. It is particularly relevant that the genetic association studies point to the role of junctional complexes and polarity since both coeliac disease and IBD are pathologically characterized by enhanced intestinal permeability[44, 45].

As discussed earlier, differentiation, barrier function, and polarity, are tightly intertwined in epithelial cells. Inflammatory signals derived from cytokines like interferon-gamma in coeliac disease[46], and both interferon-gamma and tumor necrosis factor-alpha in IBD[47], are potent in inducing impaired differentiation, enhanced permeability, and decreased polarity. We have demonstrated in this thesis (chapter 5) that arrested terminal differentiation of enterocytes is a hallmark of coeliac disease. This is less evident for IBD, but the similarity between the two disorders as mentioned above may point to a similar phenomenon in IBD. In this respect it is relevant to take note of the impaired detoxification function in both coeliac disease (chapter 5) and IBD[9]. We explained this impairment by a generic loss of specialized enterocyte functions due to an incomplete differentiation. The detoxification impairment in IBD was interpreted as a defect in transcriptional regulators (e.g. the pregnane X receptor) with an effect on multiple downstream target genes[48]. Personally, I feel it is more likely that IBD, like coeliac disease, is characterized by an inflammation-induced loss of epithelial differentiation.

In summary, coeliac disease and IBD share several characteristics: 1) they both involve the intestinal tract; 2) they co-occur in patients; 3) they are characterized by enhanced permeability; 4) they are genetically associated with the same tight junction genes; and 5) they are characterized by an impaired detoxification, probably due to a loss of differentiation. Clearly there are some differences: coeliac disease is precipitated by gluten peptides and is dominated by an HLA-DQ2 or -DQ8 mediated Th1 response, with a prominent role for *IFNG* but not *TNF*. In contrast, IBD is characterized by bacterial overgrowth partly due to defective NOD2/CARD15-mediated defensin functions[49, 50], and an inflammation driven by *IFNG* and *TNF*. In conclusion, coeliac disease and IBD may both turn out to be inflammatory epithelial de-differentiation disorders, and differ mainly in the

initiators of the inflammation (gluten versus bacteria) and the intestinal location where these present maximally (duodenum versus colon).

The Marsh story: a single pathology or 'three strikes and you're out'?
The histology of the coeliac disease mucosa is typically characterized by an infiltration with lymphocytes, both in the lamina propria and the epithelium, hyperplasia of the crypts of Lieberkühn, and atrophy of the villi. Marsh described the pathogenesis of coeliac disease by a successive increase of the initial lymphocytosis into crypt hyperplasia, ultimately leading to villous atrophy[51]. Later, the atrophic phase was further subdivided into partial, subtotal, and total villous atrophy, which particularly served diagnostic purposes[52]. This sequence of events in the pathogenesis is exactly mirrored when coeliac disease patients go on a gluten-free diet and traverse the same Marsh sequence in reverse order. Monitoring of the mucosal normalization by using the adapted Marsh nomenclature is an important tool for assessing the success of a gluten-free diet, or other treatments as in the case of refractory coeliac disease.

In chapter 3 we described the relation between the progression along the Marsh sequence and the level of inflammation as measured by *IFNG* expression, and observed a strong correlation between the two. We also noted that *IFNG* expression reaches a maximum level after which the villous atrophy still continues to progress from partial, to subtotal, to total. Apparently, progression of the villous atrophy is rather a matter of the length of time they are exposed to interferon-gamma, than a further increase in the level of this exposure. In the same study we observed that the mucosa is not a homogenous tissue, but features a mosaic of gene expression. Intriguingly, this was the case in both the normal duodenum and in the coeliac duodenum, regardless of the Marsh stage. We, and other workers, observed that in some cases limited tissue heterogeneity is also present, usually spanning no more than two consecutive Marsh stages[53-55].

This led us to propose a model in which the basic heterogeneity in gene expression, under conditions of increasing inflammation, may transform into tissue patchiness in patients. This was also based on the fact that both tissue patchiness and gene expression heterogeneity existed on a similar level of resolution (< 5-10 mm). The profile of *IFNG* expression is strongly correlated with the advance in tissue remodeling[56]. However, since not all tissue patches in the mucosa are exposed to the same level of interferon-gamma simultaneously, they do not always show the same Marsh histology. Therefore, we propose that the mucosa responds with a specific tissue remodeling only to certain ranges of interferon-gamma concentrations. We demonstrated that this tissue response to *IFNG* expression is universal and independent of the patients or their diagnosis (coeliac disease, refractory coeliac disease, or dermatitis herpetiformis). With increasing

inflammation, mucosal cells may enter the next interferon-gamma concentration interval and the tissue responds with the next stage of remodeling. However, due to the mosaic pattern of *IFNG* expression, not all tissue patches are subjected to this change simultaneously and mucosal 'patchiness' is temporarily visible.

The contrast between discrete steps in mucosal remodeling and the gradual increase in gene expression was not exclusive for *IFNG*. On the contrary, the vast majority of the gene expression profiles were characterized by a gradual increase (chapters 3 and 5) or decrease (Diosdado, submitted) during mucosal recovery. So, if we do not observe Marsh stage-specific profiles for groups of differentially expressed genes, why do we see these so markedly dissimilar histological features? It may be that the changes that drive tissue remodeling are not caused by transcriptional changes but by other modes of molecular regulation. Alternatively, there may be molecular response elements that show different levels of sensitivity to similar inflammatory signals. These elements could belong to disparate molecular pathways, programmed to execute the three major morphological changes in the coeliac mucosa. In this respect we should not consider coeliac disease as a single pathology, but rather as three consecutively mobilized pathologies. It is worth noting that some patients never progress to the stage of villous atrophy[57], or even crypt hyperplasia. Our results (chapters 3 and 5) suggest that these patients are true coeliacs, but they respond to gluten-containing food with an interferon-gamma level insufficient to traverse the entire Marsh sequence. However, due to the partial loss of enterocyte differentiation, which is proportional to the inflammation (chapters 3 and 5), they need to follow a gluten-free diet for optimal health.

Coeliac disease is a 'healthy' disease

From an evolutionary point of view, the gut is the oldest organ in multi-cellular organisms and dates back to the Precambrian era, more than 600 million years ago. Because the endoderm provides direct access to the animals' internal environment, it must have formed, from its earliest appearance, the major entrance for pathogenic and parasitic life forms. As time progressed, an evolutionary arms race between hosts and invaders will have led to the development of an integrated multi-faceted defense, consisting of innate and adaptive immunity, and cellular and physiological response mechanisms[58, 59]. Therefore, it may be fruitful to regard the pathological response to gluten peptides in coeliac disease in this evolutionary context.

A typical reaction to gluten exposure in coeliac disease patients is the physiological response with diarrhea, and the morphological change to a 'flat' intestinal mucosa devoid of villi. Diarrhea is an efficient response to flush out pathogens, and is commonly associated with bacterial infections[60]. The atrophy of the villi appears to be a well-orchestrated morphological change, leading to a

smaller surface for pathogens and parasites to attack. It is important to realize that the entire surface of the intact intestine is half the size of a soccer pitch. Together, diarrhea and a flat mucosa will result in a diminished burden, if not a total eradication, of hostile invaders. However, there is a price to pay: diarrhea can lead to severe dehydration, while villous atrophy interferes with normal intestinal activities like nutrient absorption and enterotoxin excretion. Obviously, as an acute response, this may be an efficient way to cope with hostile invaders, but in the case of chronic exposure, for example dietary gluten in coeliac disease, it puts a heavy burden on the patient's general condition.

However, there is more to this than mucosal tissue remodeling and a physiological response. In chapter 5 we demonstrated that the inflammatory response and the villous atrophy are accompanied by arrested terminal differentiation of enterocytes. The clinical effects are comparable to, and have at least the same impact as that of, a 'flat' mucosa: impaired uptake and processing of sugars, fatty acids, peptides, sterols, iron, and water, and a loss in detoxification capacity. These are the molecular changes that we have observed on the level of gene transcription. However, it is likely that similar changes could result from protein regulatory mechanisms e.g. endocytosis, compartmentalization, and post-translational modification. So, what is the possible purpose and mechanism of this impaired enterocyte differentiation? Micro-organisms like bacteria[61] and viruses[60] use cell-specific attachment sites such as membrane-bound receptors to enter cells. Likewise, proteins of tight junctions and adherence junctions are used by pathogens to breach the epithelial barrier and take the paracellular passage to enter the internal environment and the circulation[62]. The loss of differentiation, and the concomitant reduced expression or allocation of proteins from the external membrane and the junctional complex, provides an efficient defense strategy against invaders. The trade-off is an impaired intestinal function with enhanced permeability, an undesired side effect of chronic infection or food intolerance.

There is a strong correlation between impaired differentiation and villous atrophy on the one hand, and the extent of inflammation on the other (chapter 3). It was reported that stem cells in the crypts of Lieberkühn display increased mitotic activity in response to a gluten-evoked inflammation[63]. This may result in accumulation of cells in the crypts (explaining the hyperplasia), but also accelerate the migration of epithelial cells along the crypt-villus axis. Accelerated migration might not allow developing enterocytes sufficient time to fully carry out their differentiation program before they reach the villus tip and are shed into the lumen. It is more difficult to imagine that this also results in villous atrophy.

This 'accelerated conveyer belt' theory is an elegant way to explain the incomplete enterocyte differentiation and crypt hyperplasia controlled by the mitogenic property of interferon-gamma. The interferon-gamma-driven

inflammation might also activate a number of morphogenic pathways (e.g. Wnt, Notch, and Hedgehog) with an impact on villus morphology. Alternatively, we cannot rule out the possibility that impaired enterocyte differentiation is due to a partial de-differentiation program initiated by proinflammatory cytokines, but this requires further investigation.

Finally, enterocyte apoptosis and necrosis in response to proinflammatory cytokines has been reported frequently in conjunction with coeliac disease[64, 65]. Enterocytes stressed by gluten exposure express MICA and HLA-E, thereby activating cytotoxic T-cells and transforming natural killer (NK) cells into NKT-cells[64]. Activation of gluten-restricted $CD4^+$ T-cells in the lamina propria releases interferon-gamma, and could be directly involved in epithelial damage. It is known from virus infections that apoptosis is an efficient way to eradicate infected cells to stop viruses from further spreading[66]. Similarly, gluten-peptides absorbed by or diffused into enterocytes might activate such a suicide mode. It is widely accepted that innate and adaptive immune responses evoke the coeliac lesions[67]. However, lesions in the sense of holes in the epithelial layer have not been convincingly demonstrated. On the contrary, gaps in the gut epithelial layer, induced by natural cell shedding (anoikis) or by artificial damage, appear to be sealed efficiently to prevent leakage or pathogens passing across the barrier[68]. Even more improbable, in my opinion, is that villous atrophy is the result of massive apoptosis.

In conclusion, the gluten-precipitated immune response in coeliac disease sets in motion a series of events that make sense in the context of an evolutionary program developed for viral, bacterial, and parasitic containment. This emergency response leads to collateral damage that is acceptable as long as it is only temporary. The similarity between coeliac disease, tropical sprue[69], and tropical enteropathy[70] is striking, and may hold the key to understanding why, in some individuals, gluten leads to a pathological response that seems as familiar as a fight between two old enemies. In this respect it is enlightening to refer to coeliac disease as a 'healthy' disease: a proper response but against the wrong foe.

The future is now

Gene expression studies have come of age with the development of global gene sets like the Operon oligonucleotide microarray that we have used in this work, or the platforms provided by Agilent, Affymetrix and Illumina. The resolution of gene expression studies can be further enhanced by using laser capture microdissection[71], or even single-cell isolations coupled to RNA-amplification procedures[72]. However, the results described in chapter 5 show that a vast amount of information can be obtained from a single cell layer contained within a mixed tissue sample, without resorting to micromanipulation.

Genetic association studies will be dramatically accelerated by the combined availability of linkage-disequilibrium maps produced by the International HapMap Project[73], and genome-wide SNP sets (Illumina Sentrix HumanHap550; Affymetrix GeneChip® Human Mapping 500K Array Set). Complex genetic disorders may be characterized by a few genes with relatively high odds ratios (e.g. *HLA* and *MYO9B* in coeliac disease), and many more susceptibility-enhancing genes with much lower odds ratios. High-efficacy approaches will therefore require the assembly of large patient cohorts, although the potential risk of population stratification will always be lurking in the background. The near future will soon reveal the success rate and impact of new gene discoveries. However, a major problem of the results from current and future genetic association studies might be how to translate these into functional significance and patient counseling.

An interesting new development is that of genetical genomics: the application of high-throughput genetics and gene expression measurements for the identification of variants and QTL loci that control the expression of their own (cis) or other (trans) genes[74]. Processing large data streams and integrating them into functionally annotated gene networks will demand much from bioinformatics. The development of bioinformatics tools to manage and integrate the avalanche of data from genetical genomics and proteomics into comprehensive system biology will be the major challenge for the near future.

Directions for further research

The gluten-evoked inflammation in coeliac disease initiates a cascade of events that leads to mucosal remodeling and arrested enterocyte differentiation. However, how the emitted inflammatory signals influence the morphogenic program in the mucosa is largely unknown. It would be interesting to examine the role of the myofibroblast surrounding the crypts e.g. by expression profiling and Laser Capture Microscopy to identify which molecular pathways are activated. The inflammation affects mainly the terminal differentiation of enterocytes. Whether this is an indirect consequence of enhanced mitotic activity in the crypts and accelerated cell migration, or that there is a direct effect of inflammatory signals on enterocyte differentiation remains to be established. Manipulation of cell proliferation in the crypts in experimental animals by means of non-inflammatory signals might provide this information. Further insight in the differentiation gene network might be obtained by inactivation of *TM4SF4*. This gene has anti-proliferative and pro-differentiation properties and is differentially expressed in coeliac disease. Construction of cell lines with RNAi or conditional KO mice with inactivated *TM4SF4* could clarify its morphogenic role and may further facilitate identification of up- and down-stream genes in this differentiation pathway. The effects may be assessed by (immuno)histochemistry and gene expression profiling.

In addition to the arrested differentiation of enterocyte, also the loss of villi might be the consequence of increased cell proliferation and migration. Alternatively, other mechanisms should be considered including changes in the villus actin-myosin structure, induced vasoconstriction, or a pressure drop in the micro-lymph system. In addition to coeliac disease, also IBD was characterized by an impaired detoxification system, probably likewise due to an epithelial differentiation defect. It would be worthwhile to examine this in closer detail and determine which other genes and pathways may be affected and whether clinical features might be implicated. The defect in the detoxification system, both in coeliac disease and IBD, may affect the viability of the enterocytes and may also pose a risk to neoplastic transformation. Assessment of the mutation rate in such impaired epithelial cell may shed further light on this potential risk.

In studies executed parallel to this thesis we have identified the *MYO9B* gene as a susceptibility gene for coeliac disease and IBD in the Dutch population. By using a strategy of molecular pathway analysis we also identified two tight junction genes giving enhanced risk for coeliac disease and IBD. These three genes may contribute to disease susceptibility by weakening the intestinal barrier. However, since Rho-GAP domain of *MYO9B* could be involved in the cytoskeletal changes in several cell types, also other mechanisms than a barrier defect should be considered. Generation of a conditional KO mouse strain with a deleted *MYO9B* gene may reveal which cell types are affected, and what the physiological effect of this impairment may be. Likewise, an RNAi-mediated *MYO9B* knockout in cell lines may facilitate the identification of members up- and downstream in the *MYO9B* pathway by e.g. gene expression profiling. Furthermore, protein-protein interaction assays may identify new MYO9B binding proteins. Although the association studies suggest genetic involvement of tight junctions in coeliac disease (and IBD) pathogenesis, it is not clear whether the assumed permeability increase has a direct or indirect effect on disease risk. Gluten may thus find easier access via the paracellular route to the lamina propria to trigger an adaptive immune response. Alternatively, bacteria or viruses might take advantage of a genetically impaired intestinal barrier and the ensuing immune response may sensitize the mucosa for gluten. It is therefore important to determine the entrance route of gluten in the intestine.

Analysis of a complete molecular pathway to identify susceptibility genes turned out to be a successful approach in the case of the tight junction genes. A similar strategy could be adopted for other pathways, including that of the immune-synapse and –signaling, or the cytoskeleton. The advent of genome-wide SNP analyses may soon make this pathway-directed strategy obsolete and may lead to the identification of molecular pathways and at-risk genes not considered yet.

REFERENCES

1. Franke L, van Bakel H, Diosdado B, et al. TEAM: a tool for the integration of expression, and linkage and association maps. Eur J Hum Genet 2004;12:633-8.
2. Monsuur AJ, de Bakker PI, Alizadeh BZ, et al. Myosin IXB variant increases the risk of coeliac disease and points toward a primary intestinal barrier defect. Nature genetics 2005;37:1341-4.
3. Van Bodegraven AA, Curley CR, Hunt K, et al. Genetic variation in myosin IXB is associated with ulcerative colitis. Gastroenterology 2006;in press.
4. Ouellette AJ. Paneth cell alpha-defensin synthesis and function. Curr Top Microbiol Immunol 2006;306:1-25.
5. Muller CA, Autenrieth IB, Peschel A. Innate defenses of the intestinal epithelial barrier. Cell Mol Life Sci 2005;62:1297-307.
6. Lievin-Le Moal V, Servin AL. The front line of enteric host defense against unwelcome intrusion of harmful microorganisms: mucins, antimicrobial peptides, and microbiota. Clin Microbiol Rev 2006;19:315-37.
7. Berkes J, Viswanathan VK, Savkovic SD, et al. Intestinal epithelial responses to enteric pathogens: effects on the tight junction barrier, ion transport, and inflammation. Gut 2003;52:439-51.
8. Kaminsky LS, Zhang QY. The small intestine as a xenobiotic-metabolizing organ. Drug Metab Dispos 2003;31:1520-5.
9. Langmann T, Schmitz G. Loss of detoxification in inflammatory bowel disease. Nat Clin Pract Gastroenterol Hepatol 2006;3:358-9.
10. Pinto D, Clevers H. Wnt control of stem cells and differentiation in the intestinal epithelium. Exp Cell Res 2005;306:357-63.
11. Bullen TF, Forrest S, Campbell F, et al. Characterization of epithelial cell shedding from human small intestine. Lab Invest 2006;86:1052-63.
12. Sawada N, Murata M, Kikuchi K, et al. Tight junctions and human diseases. Med Electron Microsc 2003;36:147-56.
13. Schneeberger EE, Lynch RD. The tight junction: a multifunctional complex. Am J Physiol Gastrointest Liver Physiol 2004;286:C1213-28.
14. Van Itallie CM, Anderson JM. Claudins and epithelial paracellular transport. Annu Rev Physiol 2006;68:403-29.
15. Rahner C, Mitic LL, Anderson JM. Heterogeneity in expression and subcellular localization of claudins 2, 3, 4, and 5 in the rat liver, pancreas, and gut. Gastroenterology 2001;120:411-22.
16. Inai T, Sengoku A, Guan X, et al. Heterogeneity in expression and subcellular localization of tight junction proteins, claudin-10 and -15, examined by RT-PCR and immunofluorescence microscopy. Arch Histol Cytol 2005;68:349-60.
17. Fujita H, Chiba H, Yokozaki H, et al. Differential expression and subcellular localization of claudin-7, -8, -12, -13, and -15 along the mouse intestine. J Histochem Cytochem 2006;54:933-44.
18. Burridge K, Wennerberg K. Rho and Rac take center stage. Cell 2004;116:167-79.

19. Matter K, Aijaz S, Tsapara A, et al. Mammalian tight junctions in the regulation of epithelial differentiation and proliferation. Curr Opin Cell Biol 2005;17:453-8.
20. Bruewer M, Hopkins AM, Hobert ME, et al. RhoA, Rac1, and Cdc42 exert distinct effects on epithelial barrier via selective structural and biochemical modulation of junctional proteins and F-actin. Am J Physiol Gastrointest Liver Physiol 2004;287: C327-35.
21. Jaffe AB, Hall A. Rho GTPases: biochemistry and biology. Annu Rev Cell Dev Biol 2005;21:247-69.
22. Rajasekaran SA, Palmer LG, Moon SY, et al. Na,K-ATPase activity is required for formation of tight junctions, desmosomes, and induction of polarity in epithelial cells. Mol Biol Cell 2001;12:3717-32.
23. Matter K, Balda MS. Signalling to and from tight junctions. Nat Rev Genet 2003;4:225-36.
24. Wirth JA, Jensen KA, Post PL, et al. Human myosin-IXb, an unconventional myosin with a chimerin-like rho/rac GTPase-activating protein domain in its tail. J Cell Sci 1996;109 (Pt 3):653-61.
25. Alexander JS, Dayton T, Davis C, et al. Activated T-lymphocytes express occludin, a component of tight junctions. Inflammation 1998;22:573-82.
26. Zen K, Liu Y, McCall IC, et al. Neutrophil migration across tight junctions is mediated by adhesive interactions between epithelial coxsackie and adenovirus receptor and a junctional adhesion molecule-like protein on neutrophils. Mol Biol Cell 2005;16:2694-703.
27. Zen K, Parkos CA. Leukocyte-epithelial interactions. Curr Opin Cell Biol 2003;15:557-64.
28. Sung SS, Fu SM, Rose CE, Jr., et al. A major lung CD103 (alphaE)-beta7 integrin-positive epithelial dendritic cell population expressing Langerin and tight junction proteins. Journal of immunology (Baltimore, Md. 2006;176:2161-72.
29. Rescigno M, Urbano M, Valzasina B, et al. Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria. Nat Immunol 2001;2:361-7.
30. Raki M, Tollefsen S, Molberg O, et al. A unique dendritic cell subset accumulates in the coeliac lesion and efficiently activates gluten-reactive T cells. Gastroenterology 2006;131:428-38.
31. Dignass AU, Sturm A. Peptide growth factors in the intestine. European journal of gastroenterology & hepatology 2001;13:763-70.
32. Clark JA, Doelle SM, Halpern MD, et al. Intestinal barrier failure during experimental necrotizing enterocolitis: Protective effect of epidermal growth factor treatment. 2006.
33. Stepankova R, Kofronova O, Tuckova L, et al. Experimentally induced gluten enteropathy and protective effect of epidermal growth factor in artificially fed neonatal rats. J Pediatr Gastroenterol Nutr 2003;36:96-104.
34. Liu Z, Li N, Neu J. Tight junctions, leaky intestines, and pediatric diseases. Acta Paediatr 2005;94:386-93.

35. Akobeng AK, Ramanan AV, Buchan I, et al. Effect of breast feeding on risk of coeliac disease: a systematic review and meta-analysis of observational studies. Arch Dis Child 2006;91:39-43.
36. Akobeng AK, Heller RF. Assessing the population impact of low rates of breast- feeding on asthma, coeliac disease and obesity: the use of a new statistical method. 2006.
37. Cottone M, Marrone C, Casa A, et al. Familial occurrence of inflammatory bowel disease in coeliac disease. Inflamm Bowel Dis 2003;9:321-3.
38. Schedel J, Rockmann F, Bongartz T, et al. Association of Crohn's disease and latent coeliac disease: a case report and review of the literature. Int J Colorectal Dis 2005;20:376-80.
39. Tursi A, Giorgetti GM, Brandimarte G, et al. High prevalence of coeliac disease among patients affected by Crohn's disease. Inflamm Bowel Dis 2005;11:662-6.
40. Yang A, Chen Y, Scherl E, et al. Inflammatory bowel disease in patients with coeliac disease. Inflamm Bowel Dis 2005;11:528-32.
41. Curley CR, Monsuur AJ, Wapenaar MC, et al. A functional candidate screen for coeliac disease genes. 2006.
42. Stoll M, Corneliussen B, Costello CM, et al. Genetic variation in DLG5 is associated with inflammatory bowel disease. Nature genetics 2004;36:476-80.
43. Friedrichs F, Stoll M. Role of discs large homolog 5. World J Gastroenterol 2006;12:3651-6.
44. Fasano A, Shea-Donohue T. Mechanisms of disease: the role of intestinal barrier function in the pathogenesis of gastrointestinal autoimmune diseases. Nat Clin Pract Gastroenterol Hepatol 2005;2:416-22.
45. Schreiber S, Rosenstiel P, Albrecht M, et al. Genetics of Crohn disease, an archetypal inflammatory barrier disease. Nat Rev Genet 2005;6:376-88.
46. Leon F, Sanchez L, Camarero C, et al. Cytokine production by intestinal intraepithelial lymphocyte subsets in coeliac disease. Dig Dis Sci 2005;50:593-600.
47. D'Haens G. Anti-TNF therapy for Crohn's disease. Curr Pharm Des 2003;9:289-94.
48. Langmann T, Moehle C, Mauerer R, et al. Loss of detoxification in inflammatory bowel disease: dysregulation of pregnane X receptor target genes. Gastroenterology 2004;127:26-40.
49. Voss E, Wehkamp J, Wehkamp K, et al. NOD2/CARD15 mediates induction of the antimicrobial peptide human beta-defensin-2. J Biol Chem 2006;281:2005-11.
50. Begue B, Dumant C, Bambou JC, et al. Microbial induction of CARD15 expression in intestinal epithelial cells via toll-like receptor 5 triggers an antibacterial response loop. J Cell Physiol 2006;209:241-52.
51. Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('coeliac sprue'). Gastroenterology 1992;102:330-54.
52. Wahab PJ, Meijer JW, Dumitra D, et al. Coeliac disease: more than villous atrophy. Rom J Gastroenterol 2002;11:121-7.
53. Scott BB, Losowsky MS. Patchiness and duodenal-jejunal variation of the mucosal abnormality in coeliac disease and dermatitis herpetiformis. Gut 1976;17:984-92.

54. Bonamico M, Mariani P, Thanasi E, et al. Patchy villous atrophy of the duodenum in childhood coeliac disease. J Pediatr Gastroenterol Nutr 2004;38:204-7.
55. Ravelli A, Bolognini S, Gambarotti M, et al. Variability of histologic lesions in relation to biopsy site in gluten-sensitive enteropathy. The American journal of gastroenterology 2005;100:177-85.
56. Wapenaar MC, van Belzen MJ, Fransen JH, et al. The interferon gamma gene in coeliac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility. J Autoimmun 2004;23:183-90.
57. Wahab PJ. Coeliac disease: More than villous atrophy. Nijmegen, 2002.
58. Dunne DW, Cooke A. A worm's eye view of the immune system: consequences for evolution of human autoimmune disease. Nat Rev Genet 2005;5:420-6.
59. Mueller C, Macpherson AJ. Layers of mutualism with commensal bacteria protect us from intestinal inflammation. Gut 2006;55:276-84.
60. Field M. Intestinal ion transport and the pathophysiology of diarrhea. The Journal of clinical investigation 2003;111:931-43.
61. Gruenberg J, van der Goot FG. Mechanisms of pathogen entry through the endosomal compartments. Nat Rev Genet 2006;7:495-504.
62. Balkovetz DF, Katz J. Bacterial invasion by a paracellular route: divide and conquer. Microbes Infect 2003;5:613-9.
63. Kosnai I, Kuitunen P, Savilahti E, et al. Cell kinetics in the jejunal crypt epithelium in malabsorption syndrome with cow's milk protein intolerance and in coeliac disease of childhood. Gut 1980;21:1041-6.
64. Di Sabatino A, Ciccocioppo R, Cupelli F, et al. Epithelium derived interleukin 15 regulates intraepithelial lymphocyte Th1 cytokine production, cytotoxicity, and survival in coeliac disease. Gut 2006;55:469-77.
65. Maiuri L, Ciacci C, Ricciardelli I, et al. Unexpected role of surface transglutaminase type II in coeliac disease. Gastroenterology 2005;129:1400-13.
66. Koyama AH, Adachi A, Irie H. Physiological significance of apoptosis during animal virus infection. Int Rev Immunol 2003;22:341-59.
67. Londei M, Ciacci C, Ricciardelli I, et al. Gliadin as a stimulator of innate responses in coeliac disease. Mol Immunol 2005;42:913-8.
68. Watson AJ, Chu S, Sieck L, et al. Epithelial barrier function in vivo is sustained despite gaps in epithelial layers. Gastroenterology 2005;129:902-12.
69. Nath SK. Tropical sprue. Curr Gastroenterol Rep 2005;7:343-9.
70. Veitch AM, Kelly P, Zulu IS, et al. Tropical enteropathy: a T-cell-mediated crypt hyperplastic enteropathy. European journal of gastroenterology & hepatology 2001;13:1175-81.
71. Gassler N, Newrzella D, Bohm C, et al. Molecular characterisation of non-absorptive and absorptive enterocytes in human small intestine. Gut 2006;55:1084-9.
72. Bahar R, Hartmann CH, Rodriguez KA, et al. Increased cell-to-cell variation in gene expression in ageing mouse heart. Nature 2006;441:1011-4.
73. A haplotype map of the human genome. Nature 2005;437:1299-320.

74. Bystrykh L, Weersing E, Dontje B, et al. Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. Nature genetics 2005;37:225-32.

SUMMARY

Coeliac disease is a common food intolerance in Western populations, in which it has a prevalence of about 1%. In early infancy, when the transition is made to a gluten-containing diet (particularly wheat-derived products), the disease may manifest with fatty diarrhea, belly cramps, and abdominal distention. In nearly all cases these clinical symptoms disappear when patients adhere to a (life-long) gluten-free diet. However, in modern societies with a high standard of pediatric care, the proportion of patients that present with coeliac disease as an adolescent or adult is progressively growing. In these age groups coeliac disease patients show less specific clinical features like fatigue, anemia, osteoporosis, short stature, dermatitis herpetiformis (coeliac disease of the skin), and possibly reproductive failure. Whatever the clinical presentation of coeliac disease may be, it is always the result of an intestinal inflammation provoked by dietary gluten. This inflammation leads to typical changes in the intestinal mucosa: the infiltration of lymphocytes into the epithelial layer and the deeper tissue layer (lamina propria); enlargement of the crypts between the villi (crypt hyperplasia); and disappearance of the villi (villous atrophy). Thus, the intestine gets the appearance of a 'flat' mucosa. The inflammation also changes the permeability of the intestinal epithelial barrier for salt and sugar molecules (referred to as a 'leaky gut'). The loss of the villi results in a dramatic reduction of the absorptive surface of the intestine, leading to micro- and macro-nutrient deficiency that results in the various chronic health problems of untreated coeliac disease patients.

Coeliac disease is a multifactorial disorder, which means that it is caused by the interaction of environmental factors (e.g. gluten; breast feeding; infections) and heritable factors (susceptibility genes). The genetic component of coeliac disease is apparent from the enhanced risk of siblings when one sib is affected, and from comparisons between monozygous and dizygous twins. Patients with autoimmune diseases also have an increased risk of developing coeliac disease, suggesting that these diseases share common risk genes and causative mechanisms. The main genetic risk factor for developing coeliac disease is the presence of major histocompatibility complex (MHC) genes on chromosome 6p21.3 that code for the heterodimer HLA-DQ2 or HLA-DQ8. However, the presence of the MHC genes coding for these HLA molecules contribute just 40% of the genetic risk. The presence of these gene variants is necessary, but not sufficient, to develop the disease. Coeliac disease is a complex genetic disorder, meaning that multiple genes are involved in the pathogenesis, each making a minor contribution to the susceptibility. The identity of these genes, their total number, and the combinations in which they occur is, for the most part, an enigma. In this thesis we describe a combination of genetics and gene expression profiling used to identify genes and molecular pathways that contribute to the cause and pathology of coeliac disease.

Chapters 1 and 2 provide additional background information on coeliac disease in general and an overview of the various genetic approaches that have been employed. In addition, gene expression profiling is discussed as a method to gain insight into the pathology that may aid the genetic studies through the identification of molecular pathways affected by the disease. The integration of both genetics and genomics is emphasized to facilitate the identification of candidate genes and gene networks involved in the pathogenesis.

In section II the application of high-throughput technologies is presented for the study of gene expression in the duodenal mucosa of coeliac disease patients. In chapter 3 we examined the effect that tissue heterogeneity may have on the assessment of gene expression measurements. We observed that mucosal inflammation and differentiation (represented by *IFNG* and *TM4SF4* expression, respectively) were inversely correlated, and dependent on the extent of the mucosal restructuring. Interestingly, variability in gene expression between separately sampled biopsies derived from one individual was more or less independent of the variation in the histology of the same samples. This variation in gene expression between samples was present, and even comparable in size, in control samples. In other words, gene expression variation is present regardless of the pathological stage. This led us to propose a model where gene expression mosaicism in the duodenum under conditions of changing inflammation may give rise to tissue heterogeneity ('patchiness').

Chapters 4 and 5 describe the use of microarrays to study genome-wide gene expression in the duodenums of coeliac disease patients. These two studies differed in the microarray platform (cDNA versus oligonucleotides) and the type of patient samples used (case-control versus remission sequence). The two sets of differentially expressed genes thus obtained had hardly any overlap, although the biological processes they represent were comparable. The alterations induced by the pathology of coeliac disease point to an enhanced proliferation in the mucosa and an arrest in the terminal differentiation of enterocytes. This impaired differentiation particularly affected the uptake and processing of lipids, sterols, sugars, peptides, and iron. These deficiencies might contribute to the diverse clinical features seen in coeliac disease. We also observed an impaired detoxification system that was previously also described for inflammatory bowel disease (IBD). This may suggest that, like coeliac disease, IBD is also characterized by a differentiation defect.

In section III we describe how molecular pathway analysis was applied to conduct genetic (chapter 6) and gene expression studies (chapter 7). Based on the altered intestinal permeability in coeliac disease patients, and the previously identified susceptibility gene *MYO9B*, we focused on the tight junction gene network. The expression pattern of genes of the claudin family appeared to have

been evolutionary conserved for over 75 million years. Three pairs of claudins showed tight co-regulation. Patients showed variable patterns of claudin gene expression marked by extreme outliers. Particularly genes involved in signal transduction and regulation of the cytoskeleton showed trends of changed gene expression. We also observed a remarkable induction of the growth factor EGF in patients. The genetic association study of the same gene network was based on the use of single nucleotide polymorphisms (SNPs) and information on linkage disequilibrium from the International HapMap Consortium. We identified two tight junction genes that were associated with coeliac disease, and to a lesser extent with IBD. This suggests that both gastrointestinal disorders, coeliac disease and IBD share a genetic barrier defect.

The adaptive immune response provoked by gluten in coeliac disease patients is characterized by a T helper cell type 1 reaction in which interferon-gamma plays a pivotal role. In section IV in which we examine candidate genes related to the immune defense, we describe the genetics and gene expression of the *IFNG* gene. In chapter 8 we demonstrate that the extent of the mucosal remodeling is correlated with the level of *IFNG* gene expression. However, the genetic study performed with the CA-repeat from the first intron of *IFNG*, a commonly used marker for this gene, showed no genetic association in our population of Dutch patients. In chapter 9 we re-examine the *IFNG* gene, but now by using SNP markers and could demonstrate that this gene is weakly associated with coeliac disease in the Dutch population. Finally, we performed a genetic and gene expression study on four members of the *SPINK* (serine protease inhibitors of the Kazal type) gene family. *SPINK* genes are involved in tissue preservation and bacterial containment, and could therefore be considered as part of the innate immune system. Three of these *SPINK* genes mapped to two candidate regions for coeliac disease (chromosomes 5q32 and 9p21-13), which made them attractive positional and functional candidate genes. None of these *SPINK* genes, however, showed genetic involvement to coeliac disease in the Dutch population.

Finally, in the General Discussion, we evaluate our approach and our interpretations of the data, which are than incorporated in possible models of the pathogenesis of coeliac disease.

SAMENVATTING

Coeliakie is binnen de Westerse samenleving een veelvoorkomende voedingsintolerantie en heeft een prevalentie van ongeveer 1%. In de vroege kindertijd, wanneer een overgang wordt gemaakt naar glutenbevattend voedsel (m.n. tarwegerelateerde producten), manifesteert de ziekte zich middels vette diarree, buikkrampen, en een opgezette buik. In haast alle gevallen verdwijnen deze klachten als de patiënten overgaan op een glutenvrij dieet. Echter, in moderne samenlevingen die gekenmerkt worden door een hoge standaard van pediatrische zorg, is er een toenemende groei waarneembaar van patiënten die coeliakie vertonen als adolescent en volwassene. In deze leeftijdsgroepen vertonen coeliakiepatiënten minder specifieke klinische verschijnselen zoals oververmoeidheid, bloedarmoede, botontkalking, groeiachterstand, dermatitis herpetiformis (coeliakie van de huid) en mogelijk ook voortplantingsproblemen. Ongeacht de klinische presentatie van coeliakie, zij is het gevolg van een door gluten veroorzaakte ontsteking van de dunne darm. Deze ontsteking leidt tot de kenmerkende veranderingen in het slijmvlies (mucosa) van de dunne darm: infiltratie met lymfocyten in het epitheel en de dieper gelegen lamina propria; vergroting van de tussen de darmvlokken (villi) gelegen crypten (crypt hyperplasie); en het verlies van de darmvlokken (villeuze atrofie). Zo krijgt de darm het uiterlijk van een 'gladde darm". Deze ontsteking veroorzaakt ook een verhoogde doorlaatbaarheid van de darmwand voor zouten en suikermoleculen (de zgn. 'lekke darm'). Het verlies aan darmvlokken heeft een dramatische afname van het opnameoppervlakte tot gevolg en leidt tot voedingdeficiëntie in onbehandelde patiënten, resulterend in een gevarieerd beeld van chronische gezondheidsklachten.

Coeliakie is een multifactoriële ziekte, wat betekent dat het wordt veroorzaakt door een interactie van omgevingsfactoren (bijv. gluten; borstvoeding; infecties) and erfelijke factoren (risicogenen). De genetisch bedrage aan coeliakie komt duidelijk aan het licht door het verhoogde herhalingsrisico bij nakomelingen waarvan er één is aangedaan, en door het vergelijken van het risico bij eeneiige en twee-eiige tweelingen. Ook patiënten met een auto-immuunziekte hebben een verhoogde kans om coeliakie te ontwikkelen, wat duidt op gedeelde risicogenen en veroorzakende mechanismen. De belangrijkste risicofactoren om coeliakie te ontwikkelen zijn de genen gelegen binnen het MHC-complex op chromosoom 6p21.3 die coderen voor de heterodimeren HLA-DQ2 en HLA-DQ8. De aanwezigheid van de MHC-genen die coderen voor deze HLA moleculen dragen 'slechts' voor 40% bij aan het totale genetische risico. De aanwezigheid van deze genvarianten is noodzakelijk, maar niet afdoende om coeliakie te ontwikkelen. Coeliakie is een complexe genetische aandoening, wat betekent dat meerdere genen, elke met een bescheiden risicobijdrage, betrokken zijn bij het ontstaan van de ziekte. De identiteit van deze genen, hun totale aantal, en de combinatie waarin zij voorkomen is voor het grootste deel nog een raadsel. In dit proefschrift hebben we gebruik gemaakt van

een combinatie van genetische en genexpressie methoden om genen en moleculaire paden te identificeren die bijdragen aan de oorzaak en de pathologie van coeliakie.

De hoofdstukken 1 en 2 verschaffen meer achtergrondinformatie over coeliakie in het algmeeen en verder een overzicht van de diverse methoden voor genetisch onderzoek die eerder zijn toegepast. Verder wordt het vaststellen van genexpressieprofielen besproken als methode om inzicht te verkrijgen in de pathologie. Middels de identificatie van door de ziekte beïnvloede moleculaire paden kan dit een bijdrage leveren aan genetische studies. Nadruk wordt gelegd op de integratie van genetica en 'genomics' voor de identificatie van kandidaatgenen en gennetwerken die betrokken zijn bij de pathogenese.

In sectie II worden 'high-throughput' technologieën toegepast voor het bestuderen van genexpressie in de mucosa van de duodenum (twaalfvingerige darm) in coeliakiepatiënten. In hoofdstuk 3 hebben we het effect onderzocht dat weefselheterogeniteit mogelijk zou hebben op het vaststellen van genexpressie. Waargenomen werd dat de mate van ontsteking en de differentiatiegraad van de mucosa (weergegeven door de expressie van *IFNG*, respectievelijk *TM4SF4*) omgekeerd evenredig waren en afhankelijk van de mate van weefselveranderingen in de mucosa. Opmerkelijk genoeg was de variatie in genexpressie tussen afzonderlijke monsters van dezelfde persoon min of meer onafhankelijk van de variatie in weefselstructuur van dezelfde monsters. Deze variatie in genexpressie was ook aanwezig, en in de zelfde mate, in controle monsters. Dus, variaties in genexpressie zijn aanwezig ongeacht het pathologisch stadium van de darm. Dit heeft geleid tot het poneren van een model waarin een mozaïek patroon van genexpressie in de duodenum kan resulteren in weefselheterogeniteit o.i.v. veranderingen in de ontstekingsgraad.

Hoofstukken 4 en 5 beschrijven het gebruik van microarrays* voor de bestudering van genoomwijde genexpressie in de duodenum van coeliakiepatiënten. De twee studies verschilden in het type microarray (cDNA of oligonucleotiden) en het soort patiëntenmonsters (patiënt-controle of herstelreeks) dat werd gebruikt. De twee groepen van differentieel tot expressie komende genen was nauwelijks overlappend, hoewel de biologische processen die ze vertegenwoordigden wel vergelijkbaar waren. De veranderingen geïnduceerd door coeliakie wijzen op een verhoogde celdeling in de mucosa die gepaard gaat met een onvolledige ontwikkeling van de darmcellen (enterocyten). Dit ontwikkelingsdefect verstoort m.n. de opname en verwerking van vetten, sterolen, suikers, eiwit(fragmenten), en ijzer. De resulterende deficiënties dragen mogelijk bij aan het gevarieerde klinische beeld van coeliakie. Ook werd aantasting van het ontgiftingssysteem waargenomen dat reeds eerder was beschreven voor dikke

darmontstekingen (IBD). Mogelijk wordt IBD net als coeliakie gekenmerkt door een defect in de ontwikkeling van darmcellen.

In sectie III bespreken we de analyse van moleculaire paden bij de uitvoering van genetische (hoofdstuk 6) en genexpressie (hoofdstuk 7) studies. Hierbij richtten we ons op het 'tight junction' gennetwerk, dit mede gebaseerd op de verhoogde darmpermeabiliteit in coeliakiepatiënten en het eerder geïdentificeerde risicogen *MYO9B*. Het expressiepatroon van genen uit de claudin-familie bleek geconserveerd te zijn gebleven gedurende minstens 75 miljoen jaren van evolutie. Drie paren van claudin-genen vertoonden een synchrone regulatie. Patiënten lieten dynamische patronen van claudin genexpressie zien met vaak opmerkelijk afwijkende waarden. Met name genen betrokken bij signaaloverdracht en regulering van het celskelet vertoonden de tendens tot verandering in genexpressie. Patiënten gaven ook een opmerkelijke verhoging in expressie van het groeifactorgen *EGF* te zien. Hetzelfde gennetwerk werd ook onderzocht voor genetische associatie waarbij gebruik gemaakt werd van varianten van enkelvoudige nucleotiden (SNPs) en informatie over de 'linkage disequilibrium' structuur van het genoom afkomstig van het International HapMap Consortium. Twee 'tight junction' genen werden geïdentificeerd die geassocieerd bleken met coeliakie, en in mindere mate ook met dikke darmontsteking (IBD). Mogelijk delen beide darmaandoeningen, coeliakie en IBD, een genetisch defect in de darmbarrière.

De adaptieve respons van het immuunsysteem in coeliakiepatiënten, opgeroepen door gluten, wordt gekarakteriseerd door een T helpercel type 1 activatie waarbij interferon-gamma een cruciale rol speelt. In sectie IV waar we kandidaatgenen onderzoeken met betrekking tot immuniteit, beschrijven we de genetica en genexpressie van *IFNG*. In hoofdstuk 8 laten we zien dat de mate van weefselverandering in de mucosa gecorreleerd is met het niveau van *IFNG* genexpressie. Een studie uitgevoerd met de 'CA-repeat' uit het eerste intron van het *IFNG* gen (een veelgebruikte merker), vertoonde geen genetische associatie binnen onze Nederlandse patiëntengroep. In hoofdstuk 9 werd het *IFNG* gen weer getest, maar nu met SNP merkers, en werd een zwakke genetische associatie aangetoond met coeliakie binnen de Nederlandse bevolking. Tot slot hebben we ook een genetische en genexpressiestudie uitgevoerd naar vier leden uit de familie van *SPINK* (serine protease remmers van het Kazal type) genen. *SPINK*-genen zijn betrokken bij bescherming van weefsels en inperking van bacteriële groei, en kunnen worden beschouwd als onderdeel van de aangeboren immuniteit. Drie van deze vier *SPINK* genen liggen binnen twee kandidaatgebieden voor coeliakie (chromosomen 5q32 en 9p21-13), wat hun tot aantrekkelijke positionele en functionele kandidaatgenen maakte. Echter, geen van deze *SPINK*-genen bleek genetisch geassocieerd met coeliakie binnen de Nederlandse bevolking.

Ter afsluiting wordt in de Discussie een evaluatie gegeven van onze onderzoeksbenadering met interpretaties van de behaalde resultaten in de context van andere onderzoeken. Deze worden dan verwerkt in modellen m.b.t. de mogelijke oorzaak en het verloop van de ziekte coeliakie.

**) microscopisch kleine druppeltjes met het DNA van meer dan 20.000 afzonderlijke genen, aangebracht op vlakke glaasjes en geordend in een rasterpatroon)*

PUBLICATIONS

Current work

1. Diosdado B, Van Bakel H, Strengman E, Franke L, Van Oort E, Mulder CJJ, Wijmenga C, **Wapenaar MC**. A genomics view on coeliac disease points to enhanced neutrophil recruitment and barrier impairment. Clin Gastroenterol Hepatol; in press.
2. Curley CR, Monsuur AJ, **Wapenaar MC**, Rioux JD, Wijmenga C. A functional candidate screen for coeliac disease genes. Eur J Hum Genet 2006; in press.
3. Amundsen SS, Monsuur AJ, **Wapenaar MC**, Lie BA, Ek J, Gudjonsdottir AH, Ascher H, Wijmenga C, Sollid LM. Association analysis of MYO9B gene polymorphisms with coeliac disease in a Swedish/Norwegian cohort. Hum Immunol 2006;67:341-5.
4. Monsuur AJ, Stepniak D, Diosdado B, **Wapenaar MC**, Mearin ML, Koning F, Wijmenga C. Genetic and functional analysis of pyroglutamyl-peptidase I in coeliac disease. Eur J Gastroenterol Hepatol 2006;18:637-44.
5. De Haan GJ, Pinto D, Carton D, Bader A, Witte J, Peters E, van Erp G, Vandereyken W, Boezeman E, **Wapenaar MC**, Boon P, Halley D, Koeleman BP, Lindhout D. A novel splicing mutation in KCNQ2 in a multigenerational family with BFNC followed for 25 years. Epilepsia 2006;47:851-9.
6. Monsuur AJ, de Bakker PI, Alizadeh BZ, Zhernakova A, Bevova MR, Strengman E, Franke L, van't Slot R, van Belzen MJ, Lavrijsen IC, Diosdado B, Daly MJ, Mulder CJ, Mearin ML, Meijer JW, Meijer GA, van Oort E, **Wapenaar MC**, Koeleman BP, Wijmenga C. Myosin IXB variant increases the risk of coeliac disease and points toward a primary intestinal barrier defect. Nat Genet 2005;37:1341-4.
7. **Wapenaar MC**, Wijmenga C. A combined genetics and genomics approach to unravelling molecular pathways in coeliac disease. Novartis Found Symp 2005;267:113-34; discussion 134-44.
8. Diosdado B, Stepniak DT, Monsuur AJ, Franke L, **Wapenaar MC**, Mearin ML, Koning F, Wijmenga C. No genetic association of the human prolyl endopeptidase gene in the Dutch coeliac disease population. Am J Physiol Gastrointest Liver Physiol 2005;289: G495-500.
9. **Wapenaar MC**, van Belzen MJ, Fransen JH, Sarasqueta AF, Houwen RH, Meijer JW, Mulder CJ, Wijmenga C. The interferon gamma gene in coeliac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility. J Autoimmun 2004;23:183-90.
10. **Wapenaar MC**, Monsuur AJ, Wijmenga C. Erfelijke factoren bij coeliakie. Tijdschr Kindergeneeskd 2004;72:11-6
11. Diosdado B, **Wapenaar MC**, Franke L, Duran KJ, Goerres MJ, Hadithi M, Crusius JB, Meijer JW, Duggan DJ, Mulder CJ, Holstege FC, Wijmenga C. A microarray screen for novel candidate genes in coeliac disease pathogenesis. Gut 2004;53:944-51.
12. Franke L, van Bakel H, Diosdado B, van Belzen M, **Wapenaar M**, Wijmenga C. TEAM: a tool for the integration of expression, and linkage and association maps. Eur J Hum Genet 2004;12:633-8.

Previous work

13. **Wapenaar MC**, Den Dunnen JT. Exon trapping. Application of a large-insert multiple-exon-trapping system. Methods Mol Biol. 2001;175:201-15.
14. De Conciliis L, Marchitiello A, **Wapenaar MC**, Borsani G, Giglio S, Mariani M, Consalez GG, Zuffardi O, Franco B, Ballabio A, Banfi S. Characterization of Cxorf5 (71-7A), a novel human cDNA mapping to Xp22 and encoding a protein containing coiled-coil alpha-helical domains. Genomics 1998;51:243-50.
15. **Wapenaar MC**, Schiaffino MV, Bassi MT, Schaefer L, Chinault AC, Zoghbi HY, Ballabio A. A YAC-based binning strategy facilitating the rapid assembly of cosmid contigs: 1.6 Mb of overlapping cosmids in Xp22. Hum Mol Genet 1994;3:1155-61.
16. Bassi MT, Bergen AA, **Wapenaar MC**, Schiaffino MV, van Schooneveld M, Yates JR, Charles SJ, Meitinger T, Ballabio A. A submicroscopic deletion in a patient with isolated X-linked ocular albinism (OA1). Hum Mol Genet 1994;3:647-8.
17. Van Slegtenhorst MA, Bassi MT, Borsani G, **Wapenaar MC**, Ferrero GB, de Conciliis L, Rugarli EI, Grillo A, Franco B, Zoghbi HY, et al. A gene from the Xp22.3 region shares homology with voltage-gated chloride channels. Hum Mol Genet 1994;3:547-52.
18. Van De Vosse E, Booms PF, Vossen RH, **Wapenaar MC**, Van Ommen GJ, Den Dunnen JT. A CA-repeat polymorphism near DXS418 (P122). Hum Mol Genet 1993;2:2202.
19. Sieglova Z, Mrug M, Lablerova M, **Wapenaar MC**, Weissenbach J, Brdicka R. [Polymorphisms in the pseudoautosomal regions of X and Y in DNA diagnosis]. Vnitr Lek 1993;39:959-63. Czech.
20. **Wapenaar MC**, Bassi MT, Schaefer L, Grillo A, Ferrero GB, Chinault AC, Ballabio A, Zoghbi HY. The genes for X-linked ocular albinism (OA1) and microphthalmia with linear skin defects (MLS): cloning and characterization of the critical regions. Hum Mol Genet 1993;2:947-52.
21. Schaefer L, Ferrero GB, Grillo A, Bassi MT, Roth EJ, **Wapenaar MC**, van OmmenGJ, Mohandas TK, Rocchi M, Zoghbi HY, et al. A high resolution deletion map of human chromosome Xp22. Nat Genet 1993;4:272-9.
22. Meulenbelt I, **Wapenaar MC**, Patterson D, Vijg J, Uitterlinden AG. Isolation and mapping of human chromosome 21 cosmids using a probe for RTVL-H retrovirus-like elements. Genomics 1993;15:492-9.
23. Klink A, **Wapenaar M**, van Ommen GJ, Rappold G. AK1 detects a VNTR locus in the pseudoautosomal region. Hum Mol Genet 1993;2:339.
24. Bergen AA, **Wapenaar MC**, Schuurman EJ, Diergaarde PJ, Lerach H, Monaco AP, Bakker E, Bleeker-Wagemakers EM, van Ommen GJ. Detection of a new submicroscopic Norrie disease deletion interval with a novel DNA probe isolated by differential Alu PCR fingerprint cloning. Cytogenet Cell Genet 1993;62:231-5.
25. **Wapenaar MC**, Petit C, Basler E, Ballabio A, Henke A, Rappold GA, van Paassen HM, Blonden LA, van Ommen GJ. Physical mapping of 14 new DNA markers isolated from the human distal Xp region. Genomics 1992;13:167-75.
26. Breuning MH, Saris JJ, Dauwerse JG, Breslau-Siderius L, **Wapenaar MC**, van Ommen GJ. Search for the gene responsible for polycystic kidney disease and its clinical consequences. Adv Nephrol Necker Hosp 1992;21:117-24.

27. Driesen MS, Dauwerse JG, **Wapenaar MC**, Meershoek EJ, Mollevanger P, Chen KL, Fischbeck KH, van Ommen GJ. Generation and fluorescent in situ hybridization mapping of yeast artificial chromosomes of 1p, 17p, 17q, and 19q from a hybrid cell line by high-density screening of an amplified library. Genomics 1991;11:1079-87.
28. Henke A, **Wapenaar M**, van Ommen GJ, Maraschio P, Camerino G, Rappold G. Deletions within the pseudoautosomal region help map three new markers and indicate a possible role of this region in linear growth. Am J Hum Genet 1991;49:811-9.
29. **Wapenaar MC**, Pearson PL, van Ommen GJ. P9 (DXYS75) detects a VNTR-type RFLP in the pseudoautosomal region. Nucleic Acids Res 1990;18:384.
30. Kievits T, Devilee P, Wiegant J, **Wapenaar MC**, Cornelisse CJ, van Ommen GJ, Pearson PL. Direct nonradioactive in situ hybridization of somatic cell hybrid DNA to human lymphocyte chromosomes. Cytometry 1990;11:105-9.
31. **Wapenaar MC**, Kievits T, Meera Khan P, Pearson PL, Van Ommen GJ. Isolation and characterization of cell hybrids containing human Xp-chromosome fragments. Cytogenet Cell Genet 1990;54:10-4.
32. Wulff K, Herrmann FH, **Wapenaar MC**, Wehnert M. Deletion screening in patients with Duchenne muscular dystrophy. J Neurol 1989;236:470-3.
33. Sugino S, Fujishita S, Kamimura N, Matsumoto T, **Wapenaar MC**, Deng HX, Shibuya N, Miike T, Niikawa N. Molecular-genetic study of Duchenne and Becker muscular dystrophies: deletion analyses of 45 Japanese patients and segregation analyses in their families with RFLPs based on the data from normal Japanese females. Am J Med Genet 1989;34:555-61.
34. Den Dunnen JT, Grootscholten PM, Bakker E, Blonden LA, Ginjaar HB, **Wapenaar MC**, van Paassen HM, van Broeckhoven C, Pearson PL, van Ommen GJ. Topography of the Duchenne muscular dystrophy (DMD) gene: FIGE and cDNA analysis of 194 cases reveals 115 deletions and 13 duplications. Am J Hum Genet 1989;45:835-47.
35. Blonden LA, den Dunnen JT, van Paassen HM, **Wapenaar MC**, Grootscholten PM, Ginjaar HB, Bakker E, Pearson PL, van Ommen GJ. High resolution deletion breakpoint mapping in the DMD gene by whole cosmid hybridization. Nucleic Acids Res 1989;17:5611-21.
36. Robertson ME, **Wapenaar MC**, Malcolm S. A new polymorphism for the dystrophin intragenic probe P20 [DXS269] using BstX1. Nucleic Acids Res 1989;17:3327.
37. Hart KA, Abbs S, **Wapenaar MC**, Cole CG, Hodgson SV, Bobrow M. Molecular deletions in the Duchenne/Becker muscular dystrophy gene. Clin Genet 1989;35:251-60.
38. **Wapenaar MC**, Kievits T, Hart KA, Abbs S, Blonden LA, den Dunnen JT, Grootscholten PM, Bakker E, Verellen-Dumoulin C, Bobrow M, et al. A deletion hot spot in the Duchenne muscular dystrophy gene. Genomics 1988;2:101-8.
39. Bartlett RJ, Pericak-Vance MA, Koh J, Yamaoka LH, Chen JC, Hung WY, Speer MC, **Wapenaar MC**, Van Ommen GJ, Bakker E, et al. Duchenne muscular dystrophy: high frequency of deletions. Neurology 1988;38:1-4.

40. Hofker MH, Skraastad MI, Bergen AA, **Wapenaar MC**, Bakker E, Millington-Ward A, van Ommen GJ, Pearson PL. The X chromosome shows less genetic variation at restriction sites than the autosomes. Am J Hum Genet 1986;39:438-51.
41. Hofker MH, **Wapenaar MC**, Goor N, Bakker E, van Ommen GJ, Pearson PL. Isolation of probes detecting restriction fragment length polymorphisms from X chromosome-specific libraries: potential use for diagnosis of Duchenne muscular dystrophy. Hum Genet 1985;70:148-56.

ACKNOWLEDGEMENTS

The development of one's personal style in conducting research is like that of a multifactorial trait: it's partly nature and partly nurture. For transmission of my 'curiosity' genes I have to thank my parents and their ancestors. For my scientific nurture I am much indebted to the inspiring tutors that guided me while traveling the crooked path of scientific discovery. These were (in chronological order) the professors Peter Pearson, Norio Niikawa, Gertjan van Ommen, Andrea Ballabio, and Cisca Wijmenga.

Peter, you were my first tutor and always impressed me with your sharp analyses and getting to the core of an issue. Regardless of whether the subject matter is part of your expertise or not, you always know how to give its presenter a hard time. No better way to learn debating skills and how to defend one's point of view than working for you. In those pioneering days of DNA research we were a closely knit group that knew how to match hard work with high-standard entertainment: renowned were our mid-winter theme parties! My partners in crime were: Tim (my first technician), Marten (my first 'boss'), Bert (Mr. Fix-it), Peter-guitar-D, Eline (Pientje), Martijn, Mette, and Arthur.

Norio, spending five months in your lab in Nagasaki has left an everlasting impression, and a deep affection for Japan. I am grateful that you shared with me your passion for its history and archeology: we talked for hours, sometimes till early morning. I am grateful for the lecture tour you planned that took me around the southern part of Japan. My stay in Nagasaki would not have been the same were it not for your students Junichi, Shigeto, Hidefumi, Kamei (Mr. Turtle), and Tadashi. Junichi, thanks for using your apartment, and introducing me to sensei Morita: I vividly remember my colorful and swollen hand after crushing my first roof tile. Thank you guys, for introducing me to temple sites, nightclubs, historic places, tea ceremonies, and the hospitality of your families.

Gertjan, from you I adopted the custom of using figurative speech to clarify complex matters. Your use of a less formal writing style was also very instructive. Although we differ in various character aspects, we do share a passion for the experimental detail. From you I also learned that there may be 99 solutions for solving a single problem. That this renders the person asking your advice in a state of total confusion is just part of the educational experience, comparable to the teachings of Zen masters. With great fondness I think of our small YAC-team, Marijke, Annemieke, and Kurt ('*YAC clowning is fun*' [sic]), and how we set our teeth into this pièce de résistance. Other remarkable characters in those days were: Ieke-the frozen leg-Ginjaar, Johan, Lau ('*this hole here you cannot see, 'cause it's not there*'), Disco-Petra, Marleen, Nicole, Riccardo, and Monique. Thank you all for a memorable time.

Andrea, from you I learned that doing five times more experiments than necessary yields the required result (the 'bulldozer'). Also, distributing rumors

about results (that you have not achieved, yet) is a good practice to discourage competitors. It is 'extremely important' to design your envisaged publication before you start any experiments, and apply for a grant when the research is nearly done. I'd like to thank my fellow chromosome Xp gene-and-disease hunters for an exciting time in Houston: Marjon (my student), Giovanni Battista ('*macellaio*'), Elena ('*sono stanco*'), Antonella, Giuseppe, Barbara, Sandra, Brunella, Maria Teresa, and Brett; and from Huda Zoghbi's lab: Linda and Sandro ('*Hello, I am not available*').

Cisca, when I was almost on the brink of abandoning science I was asked to join your team, in an original manner as only Peter Pearson can do. It was a tough transition going from physical mapping and cloning to the unfamiliar field of complex genetics, immunology, and epithelial biology. It has been a steep learning curve and, at times, I felt a bit lost. However, despite these difficulties, it has been a most fascinating and rewarding challenge. Although we have different styles of working, we match perfectly when it comes to scientific analysis and planning. I do appreciate the way you welcome new ideas and transform them into executable plans. We also share a special attitude towards deadlines, although we differ slightly in the details.

I am grateful for contributions from the following external collaborators, without whom this research would not have been possible: Prof. Chris Mulder, Dr. Luisa Mearin, Dr. Joachim Schweizer, Dr. Jos Meijer, Prof. Gerrit Meijer, Prof. Frank Holstege, Drs. Mohammed Hadithi, Dr. Roderick Houwen, and Dr. David van Heel with colleagues. Thanks to the participants in the CDC consortium for fruitful discussions. I am particularly grateful for the participation of the coeliac disease patients with their families, and the NCV. Their help is indispensable.

I thank my students Justin, Jos, and Mieke, whose enthusiasm and productivity was most welcome. I sincerely thank Begoña for the microarray hybridizations, Harm and Lude for help with the bioinformatics, Alienke, Sasha, and Martine for the genetic studies, and the other members of the coeliac disease team for their involvement: Erica, Eric, Marianna, Ruben, and Gosia. Jackie, I am grateful for all the time you've spent on revising and editing our manuscripts that, without doubt, contributes to their acceptance for publication. Flip and Harry, you have both spent countless hours on structuring and feeding the DNA database, without this it would be impossible to even start thinking of conducting genetic research. I would also like to thank all the other members of our Complex Genetics Group for creating a productive and pleasant environment. Dalila and Behrooz, thanks for always being available for a late evening chat or brainstorm. Judith, thanks for keeping us all in tune.

My dear friends André, Edith, Adri, Karin, and Eline. Thanks for your unabated faith and moral support that I would eventually complete this thesis. I hope that we will soon have time to enjoy what we like to do most: making

pretentious photo expeditions in well-catered nature reserves; enjoying a classical concert; judging an expensive wine on its qualities while having a good meal; and, of course, having a good laugh. Antonella, Ugo, Carlo, and Veronica, you have been, and still are, my second family in Naples. We went through a great deal while Lisa was ill, and later after she passed away. I cannot imagine where I would have been without your support. This thesis is dedicated to Lisa, and I am sure she would have been pleased with its contents, which extend beyond mere genome mapping. I would also like to dedicate this thesis to my parents; they died before they could witness its completion. I am particularly grateful to my father who, without any overstrained anticipation, always supported me in my aspirations and ignored the patronizing advice from my school teachers. I dreamt of being a scientist from an early age, without any notion of what it really meant, or the discipline I would choose. With worse than mediocre results at primary school, this was a joke in the eyes of many, not least my teachers. However, persevering in this dream and the love of discovering the unknown has brought me, be it with ups and downs, to this point. Without this inner drive I would have abandoned science a long time ago. To any young wannabe scientist, I would say they should just follow their inner compass: '*It don't mean a thing if it ain't got that swing*'.

CURRICULUM VITAE

The author of this thesis was born on 14 July 1958 in Vlaardingen, The Netherlands. After graduating high school, HAVO (1975) and VWO (1977), at Scholengemeenschap Westland-Zuid in Vlaardingen he entered in 1978 the Biology program at Leiden University, The Netherlands. In 1984 he graduated *cum laude* for his Master of Science degree with a Major in Molecular Genetics. As part of his Master's training he conducted 24 months of experimental research during three internships. The first seven months were spent on the 'Evaluation of methods for isolating nuclei from undifferentiated tissue cultures from the tobacco plant' at the Department of Botanical Morphogenesis (Prof. Dr. Libbenga), at Leiden University. The following twelve months were devoted to the 'Isolation and characterization of single copy DNA markers from flow-sorted X chromosome-specific phage lambda libraries' at the Department of Human Genetics (Prof. Dr. P.L. Pearson), at Leiden University. Finally, for five months he worked on the 'Construction of a cDNA library from rat liver to study aging, using differential gene expression' at the Institute for Experimental Gerontology (Prof. Dr. D.L. Knook), IVEG-TNO, The Netherlands. After his graduation in 1984 he enrolled a PhD program at the Department of Human Genetics (Prof. Dr. P.L. Pearson), Leiden University, on the NWO project 'Development of methods for walking along human chromosomes'. During this period (three and a half years) he worked for one month on 'chromosome-mediated gene transfer' at the Human Molecular Genetics Laboratory (Prof. Dr. P. Goodfellow), ICRF, London. In the period January – May 1988 he was a Visiting Research Fellow at the Department of Human Genetics (Prof. Dr. N. Niikawa), at the Nagasaki University School of Medicine, Japan. During this period he visited several academic centers in southern Japan where he gave lectures on the 'genetics of Duchenne muscular dystrophy', and 'DNA fluorescent *in situ* hybridization'. Next, he returned to the Department of Human Genetics (Prof. Dr. G.J.B. van Ommen) at Leiden University where he worked as Research Assistant on the construction of human DNA libraries using Yeast Artificial Chromosome (YAC) vectors. From February 1992 till September 1994 he was attached as Postdoctoral Fellow to the Institute for Molecular and Human Genetics (Prof. Dr. A. Ballabio), at Baylor College of Medicine in Houston (TX), USA. Here he worked on cloning and mapping disease genes from the distal portion of human chromosome Xp, and dedicated himself to the MLS syndrome. After a career break he returned in August 1998 to the Department of Human and Clinical Genetics at Leiden University. He worked, initially as volunteer and later as Research Assistant, with Dr. J.T. den Dunnen on the EC project 'Development of systematic exon trapping for gene identification in genomic DNA' for the development of the method of 'cosmid-based exon trapping'. Since January 2001 the author is employed as Postdoctoral Fellow in the Complex Genetics Group of Prof. Dr. C. Wijmenga at the DBG-Department of Medical Genetics, University Medical Center Utrecht, The Netherlands. It is here where he works on the genetics and genomics of coeliac disease and where this thesis was written.

COLOR ILLUSTRATIONS

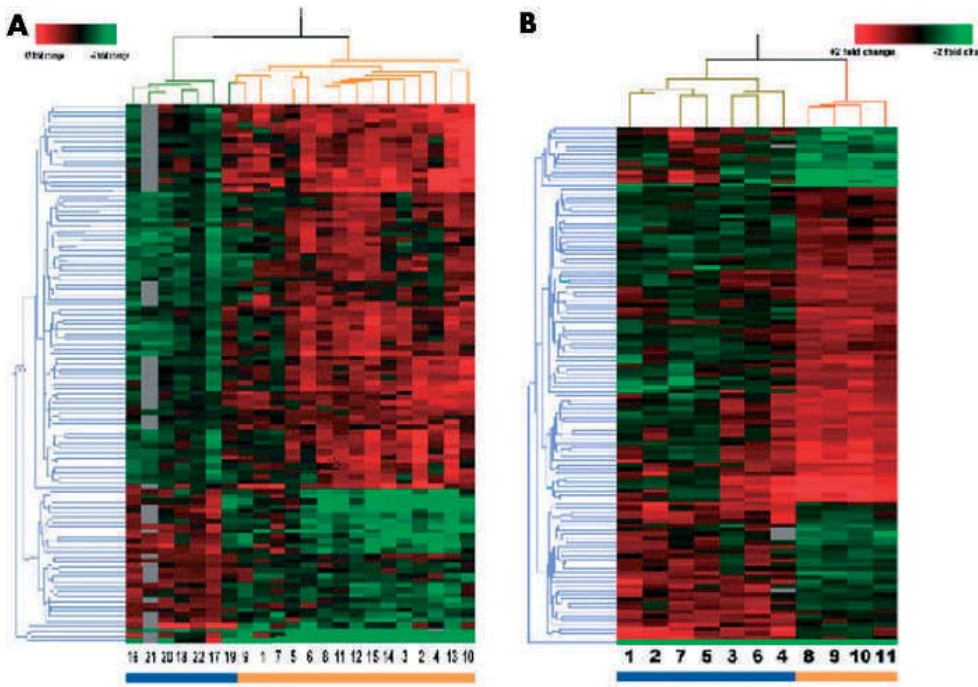


Fig. 1, page 68

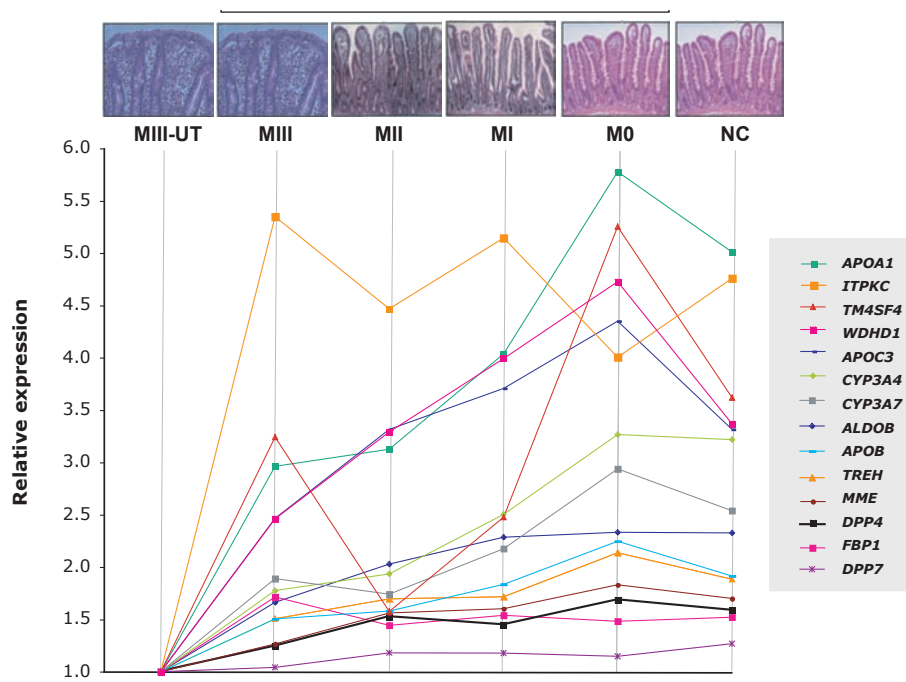


Fig. 1, page 78

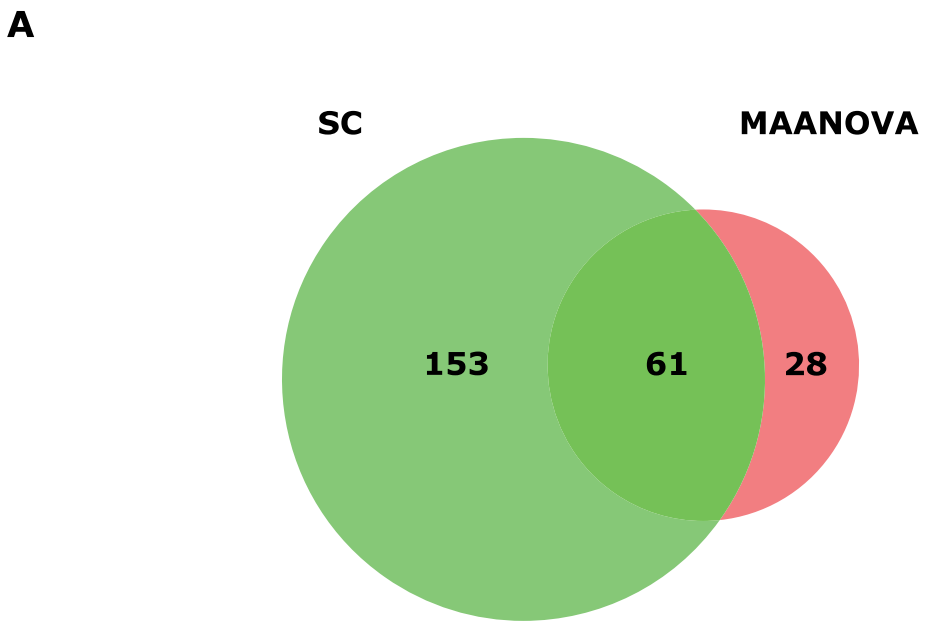


Fig. 2A, page 79

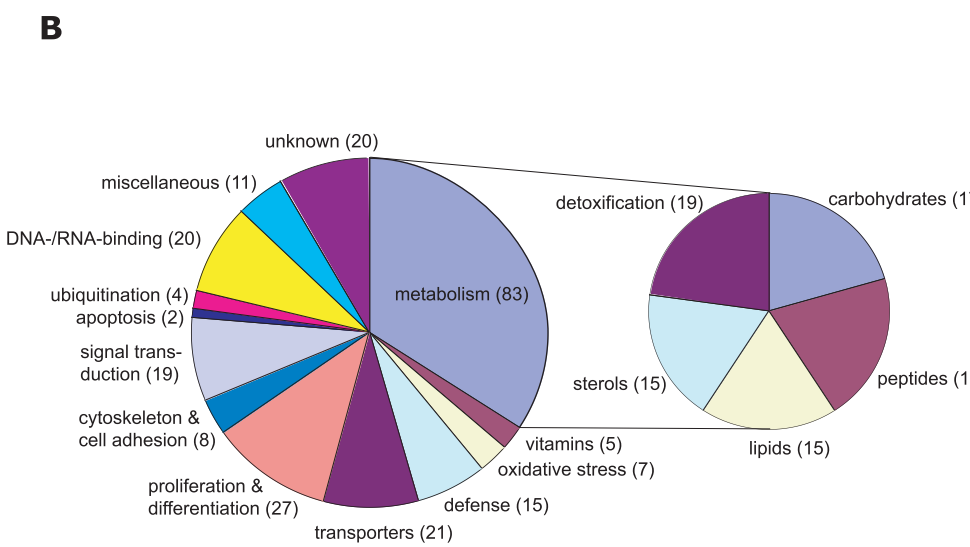


Fig. 2B, page 79

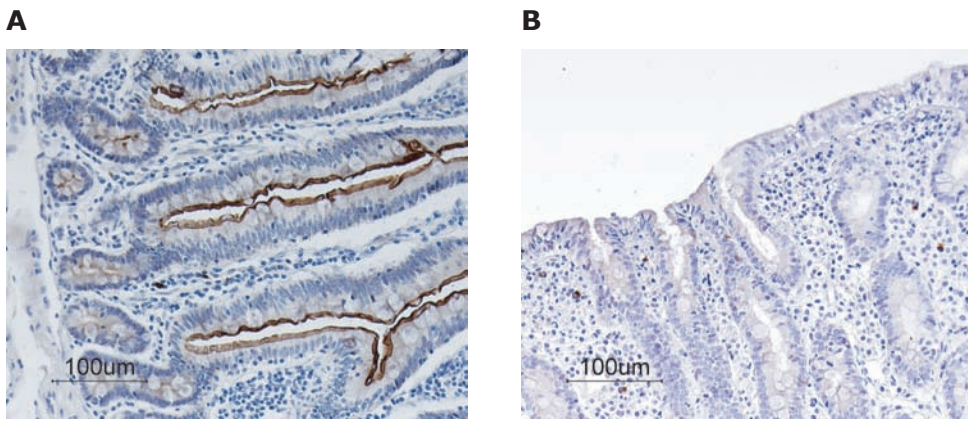


Fig. 3, page 89

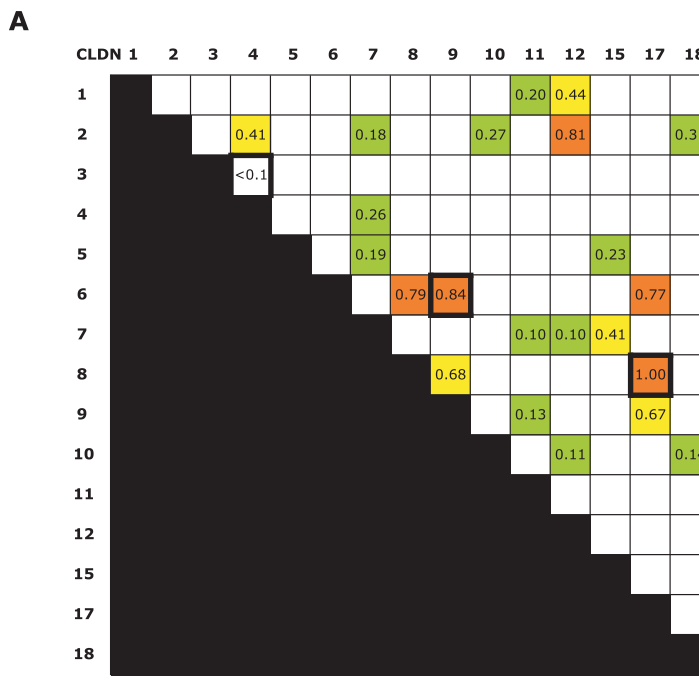


Fig. 3A, page 114

B

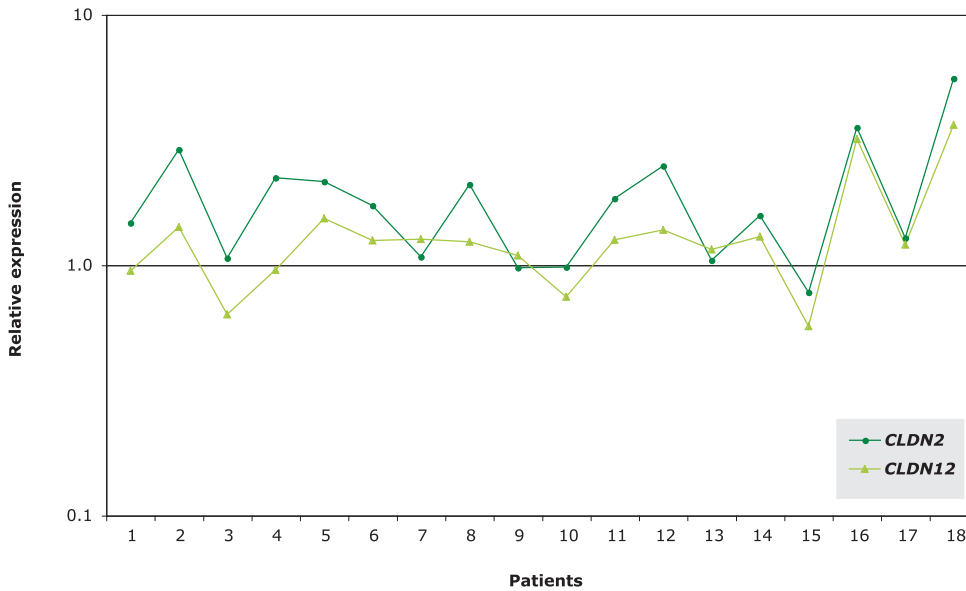


Fig. 3B, page 115

C

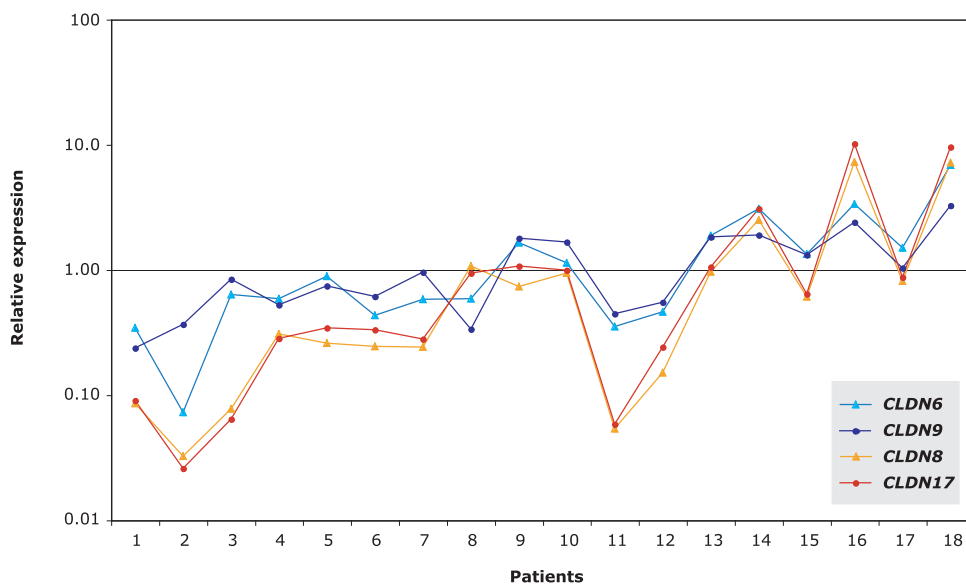


Fig. 3C, page 115

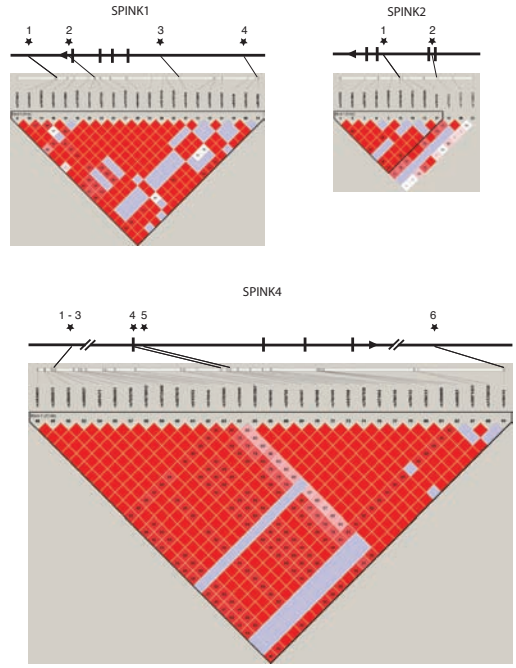


Fig. 2, page 162

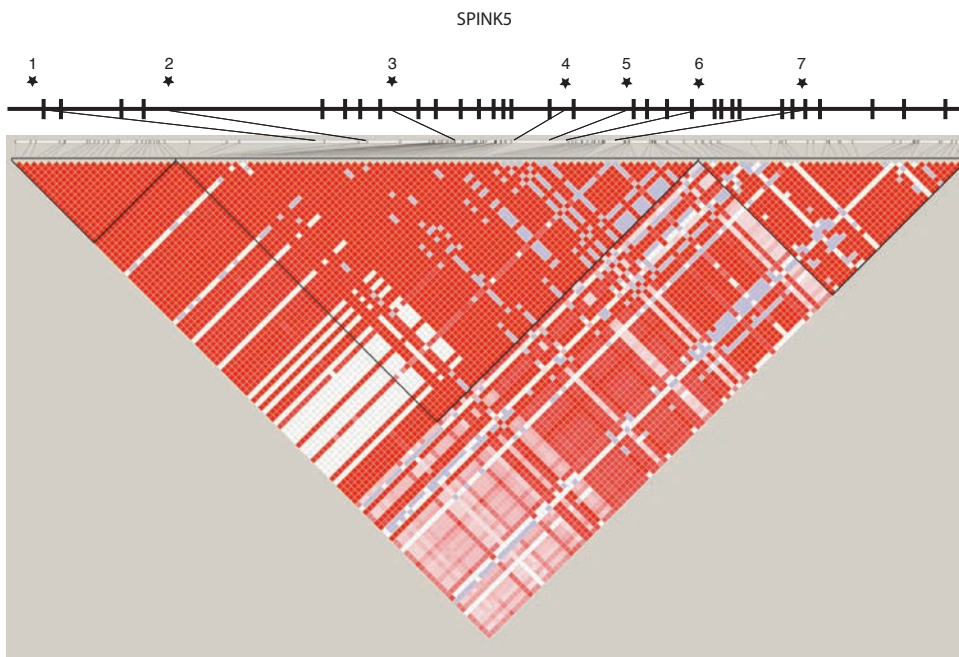


Fig. 2, page 163